

~~Dear team, I hope this finds you well~~
~~We are pleased to share I owe you an apology~~
~~for the delay. Due to unforeseen~~
~~The report took longer because I rewrote it~~
~~twice three times. Section 2 still needs you.~~
too formal? say it plainly
~~Thank you for waiting. Best regards D.~~

TRUST AND AUTHENTICITY IN AI-ASSISTED
PROFESSIONAL COMMUNICATION

Make It Sound Human

*What Readers Look For When Machines Help
Us Write*

WILLIAM DEKKER

Make It Sound Human

TRUST AND AUTHENTICITY IN AI-ASSISTED PROFESSIONAL
COMMUNICATION

Make It Sound Human

*What Readers Look For When Machines Help Us
Write*

WILLIAM DEKKER

Copyright © 2026 William Dekker. All rights reserved.

No part of this book may be reproduced in any form without written permission from the author, except for brief quotations in reviews and scholarly work.

This book is adapted from the author's doctoral research at the School of Communication, Daystar University, Nairobi. The participants are identified by code only, and names that appeared in their words have been masked.

The workplace scenario described in this book, including the agency, its client and the people named in the two messages, is invented. Any resemblance to real organisations or people is coincidental.

ISBN 978-9914-29-663-1 (paperback)

ISBN 978-9914-29-664-8 (ebook)

Published by William Dekker, Nairobi, Kenya

*To my parents, who took me to school with every penny in its
scarcity, and to my teachers, who taught me to read and write, as
here, in return, I do.*

Contents

Author's Note	ix
Introduction: Did a Person or Machine Write This?	1
Chapter One: The Hand in the Text	11
Chapter Two: The Ledger of Effort	25
Chapter Three: Written to Me, or Written at Me	37
Chapter Four: When Authenticity Matters, and When It Does Not	49
Chapter Five: Prosecuting From the Dock	59
Chapter Six: A Warrant With an Expiry Date	71
Chapter Seven: What Writers, Teams and Tool Makers Can Do	83
Checklist: Before You Hit Send	95
Appendix: How This Study Was Done	97
References	107
Acknowledgements	117
About the Author	119

Author's Note

This book grew out of my doctoral research at Daystar University in Nairobi. The title comes from a market that has grown up around AI writing tools: products whose whole purpose is to make machine-written text sound more human. If such products sell, readers must be noticing something. I wanted to find out what.

Sixteen professionals working in Nairobi helped me answer that question. They read two versions of the same workplace email, one written by me and one by ChatGPT from the same facts, and talked about what they noticed and what they trusted. Only afterwards did they learn which was which. This book is built on what they said. I have quoted them exactly as they spoke, hesitations included, because the way people reached for words was often part of what they were telling me. Where I have left out part of what someone said, the gap is marked [...].

The participants appear as P1 to P16. I promised them anonymity, so nothing here links a quotation to a person, an organisation or a role.

Turning a dissertation into a book meant changing its order and its voice. The six themes from the analysis now give the book its chapters, and each chapter ends with what the finding means for anyone who writes at work. The literature review and the methods have been folded into the chapters, and a plain account of how the study was done, with both emails in full, is at the back. The “In Practice” notes in each chapter draw on the study and on wider research into machine-written text.

Sixteen people in one city may not tell us how common these views are, but they can show, in detail, how a reader makes sense of a message they suspect a machine wrote, and what a writer can do about it.

William Dekker
Nairobi, 2026



INTRODUCTION

Did a Person or Machine Write This?

A Message About a Lost Account

One afternoon, just before a weekend, a team leader at a communications agency in Nairobi learns that the agency has lost its largest client. The Sankara account has gone to a competitor, and the decision came down to price. The team worked hard on the pitch. Someone now has to give them the bad news, look after them, and point them to the next step, a team meeting on Monday.

I wrote that message twice, in a sense. The first version I wrote myself, as I would have written it in my own work as a communications specialist. It opens with “Good afternoon team,” and it is short. It says the agency has heard from Sankara that its bid to keep the account has failed, and that the work is going to another agency. It names a colleague, Asha, whose pitch was strong. It says the full feedback has been requested and will be shared. It proposes a meeting on Monday at 9, wishes everyone a restful weekend, and is signed with a first name only: “Best, Siddondo.” It runs to 147 words, including the greeting and the sign-off.

The second version came from ChatGPT, a tool many readers of this book will have used themselves. I gave it a prompt carrying the same facts: the account lost, the price, the strong pitch, the Monday meeting. I asked it to keep the message professional. It opened with “Dear Team,” and its first sentence read: “I am writing to inform you that we have received confirmation this afternoon that the Sankara account has been awarded to another agency.” It thanked the team for their “hard work, commitment, and professionalism”, told them that “While we cannot change the outcome, we can control how we respond,” encouraged everyone to remain positive and focused, and closed with “Kind regards, Siddondo P. A.” It ran to 222 words, half as long again as mine, in five body paragraphs where mine had three. Somewhere in the third paragraph the tool had placed a long dash.

Both messages are competent, and neither contains an error of fact. Yet when sixteen working professionals in Nairobi read the two side by side, without being told which was which, most of them soon formed a view about which one a person had written and which one a machine had produced, and the groups largely agreed. Some of them could say why. Others knew before they could explain it.

That second experience, knowing before explaining, is where this book begins. One participant described it from ordinary working life, before the two messages had even appeared on the screen: a colleague’s email arrives, and before the reading is properly under way, you already doubt that the colleague wrote it.

Most people who work with email will recognise that moment. You may not be able to point to the sentence that gave it away, and you may be mistaken. But the question has formed, and it is a new question for most of us at work: who wrote this?

Why the Question Now Matters at Work

For most of the history of professional writing, the question had an easy answer. Senior people have long had speechwriters, and I have been one.

But the reader could assume that a person who knew the people being written to had chosen the words. That assumption has weakened quickly. By 2024, three in four knowledge workers around the world reported using AI in their daily work, and writing help was among its most common uses.¹

The tools write well, by most measures. In a study of 887 working professionals, Coman and Cardon found that AI-generated workplace messages were rated as professional, effective, efficient, confident and direct.² In a related study, however, Cardon and Coman found that only 40 to 52 per cent of professionals saw their supervisors as sincere when those supervisors used a high level of AI assistance, against 83 per cent when the assistance was low.³ Yet readers are poor at spotting machine text: in experiments with 4,600 participants, Jakesch and colleagues found that people relied on cues the machine produced just as readily.⁴ Competence holds, and something to do with sincerity and care is at risk.

Studies like these record a rating on a scale. They cannot tell us what the moment of doubt is like for the person reading, what they notice, what they decide about the sender, or what they do next. Managers, communications teams and HR departments are already deciding when to use these tools with little evidence of what those decisions cost them with the people they write to. A message that is believed but not trusted still does some of its work. It may not do the part that matters most when the news is bad.

The study behind this book asked readers directly. Its purpose was to explore how people make sense of professional messages they believe were written by AI, and how that belief shapes their sense of whether the message is authentic and whether the sender can be trusted. I chose the city with care.

Why Kenya, and Why Nairobi

Kenya is a good place to study how readers respond to messages they think AI wrote. In July 2025, a global digital report found that 42.1 per cent of

Kenyan internet users aged 16 and over had used ChatGPT in the previous month, which placed Kenya first in the world.⁵ The same year, the government published a National AI Strategy for 2025 to 2030 that names trust as a foundational goal, without saying what that trust means in practice.⁶

Nairobi is Kenya's capital and its main centre of corporate, government, non-governmental and media work. A professional there may draft a report in formal English, discuss it with colleagues in a mixture of English and Kiswahili, and sign off a quick message in a register that only people in that office would recognise as theirs. Office English in Nairobi has habits, courtesies and borrowings of its own, and a reader who knows them can hear where a message comes from.

This matters because a language model writes by predicting likely next words, and so leans towards the most common choices. In an experiment with Indian and American writers, Agarwal and colleagues found that AI writing suggestions moved people's writing towards Western styles and reduced the markers of their own cultures.⁷ In the dissertation I called this pull linguistic mean reversion: the flattening of an individual voice towards a statistical average. For a Nairobi reader, that average is not neutral. It is weighted towards the writing that dominates the tools' training material, which is overwhelmingly English and largely drawn from Western internet platforms.

African communication scholars have long argued that theories built elsewhere fit this continent poorly.⁸ Langmia and Haddad Nunez describe African communicative practice as placing weight on communal interaction, on accountability to the people addressed, and on the presence of the speaker in what is said.⁹ On that reading, a person's voice shows that they are present, that they answer to their audience, and that they think the audience worth their attention. I use this scholarship as a lens on Nairobi workplaces. It does not describe how everyone on the continent communicates, and the study treated whether it held for these readers as something to find out.

Abdulhamid and colleagues found small business owners and staff in Kenya and Nigeria describing editing generative AI output so that it would sound like them, and said the tools carried a Western bias that did not fit their customers or their way of speaking.¹⁰ Those were writers talking about their own text. Nobody, as far as I could find, had asked Kenyan readers what they experience when a message they receive seems to have come from a machine.

The risk is a flattened global register: messages that are fluent, polite and correct, and that could have been sent from any office in any country by anyone. For a reader whose sense of a real person in a text was formed in a different tradition, that register may remove the signs they look for. I held this as a possibility to examine, and the study makes no claim that African readers feel the loss more sharply than anyone else, because no comparison was made. One participant, answering a question I put to the group, said they would miss “that African touch” if it went, and Chapter 1 reports such accounts with the care they need.

Cues That Are Hard to Fake

To think about what a reader is looking for, I drew on the study of online communication. Joseph Walther spent much of his career showing that people form impressions of each other from text alone. Without faces, voices or gestures, readers pay closer attention to whatever is left: word choice, sentence habits, courtesy, the small features of how someone writes.¹¹

Online, though, people control how they present themselves, so how does a reader decide what to believe? Walther and Malcolm Parks proposed what they called the warranting principle: readers give more weight to information the person could not easily have manipulated, and less to information they could.¹² A claim such as “I care deeply about this project” costs nothing to write, and readers discount it. David DeAndrea later sharpened the idea: the harder a piece of information is to fake, the more

it is trusted, and the difficulty has two parts, the effort it would take and the access it would require.¹³

The principle was developed for social media profiles, dating sites and the like, where the question is whether a person is presenting themselves truthfully. AI writing adds an earlier question. Before a reader can ask whether the sender is sincere, they now have to ask whether the sender wrote the message at all. The study applied the principle to that question, and proposed an extension of it.

The extension is that the small, involuntary habits of a person's writing work as a warrant of authorship. Everyone who writes regularly has them: a favourite phrase, a way of opening, a punctuation habit, a tendency to keep sentences short, a turn of local idiom. These habits are hard to fake on both of DeAndrea's counts. They take effort to imitate, and they depend on access to something only that writer has, which is their own history with language. When a reader finds those habits in a message, they have evidence that a particular person wrote it. A message that is fluent, correct and unremarkable may still be accurate, but it offers nothing that connects it to anyone.

Two qualifications came with the extension, and both turned out to matter. A habit can only work as a warrant for a reader who knows it: if you have never received an email from me, my habits mean nothing to you. For such readers, I proposed, the texture of a text might still signal that some person, as opposed to a process, wrote it. And the signs of a real person are learned within a culture, so a warrant only works for a reader equipped to read it.

The study also drew on an older expectation, which I called the communicative effort contract. The philosopher Paul Grice observed that people in conversation assume a shared effort at cooperation, and Sperber and Wilson added that when someone asks for our attention, they implicitly promise that our effort in reading will be repaid.¹⁴¹⁵ The effort contract is my label for that expectation at work: that a sender has put in effort in proportion to the attention they ask for, and that this effort is a

sign of regard. It turned out to describe a good deal of what participants said.

FIGURE 1 What a reader can check

EASIER TO FAKE → HARDER TO FAKE

EASY FOR ANY WRITER, OR ANY TOOL, TO PRODUCE

A claim about yourself: “I care deeply about this team”

A warm greeting and a polite sign-off

Smooth, correct, well-organised prose

HARD TO PRODUCE UNLESS YOU ARE THE PERSON

A colleague’s name and what they did

Your own habits and slips: the dots you always use, a word dropped in haste

A fact about someone’s life that only those close to them would know

Readers give more weight to cues the sender would find hard to fake. The examples on the right are the ones participants in this study relied on.

How the Study Worked

The study was qualitative, aiming to understand experience rather than count it. Sixteen professionals took part, all based in Nairobi, each with at least three years of professional, academic or institutional experience, each reading professional communication in English as part of their work, and none working as an AI researcher or developer. They came from private, public, intergovernmental, non-profit, faith-based, academic, professional services and health organisations. Nine were men and seven were women. They appear only as P1 to P16, and I do not link any code to a job, sector or organisation, because in a sample this size those details could identify someone.

They met in four online focus groups of four people each, on Microsoft Teams, over four evenings in July 2026. I moderated every session, and each ran for about an hour and a half. Participants first talked about their everyday experience of professional messages and AI writing tools, then read the two Sankara messages, unlabelled, and responded. Only after their first judgements did I tell them which message I had written

and which came from ChatGPT, and we then explored whether that changed their reading.

The two messages share the scenario and the core facts, but they differ in more than authorship: length, structure, formality, a named colleague, a promise of feedback, a first-name signature. I did not control those differences, so participants' reactions cannot be read as a clean measure of the effect of machine authorship. They are better read as what happens when experienced readers compare two plausible ways of writing the same hard message. Three groups saw my version first and one saw the machine's first, too few to say anything about order. I stopped at four groups because the fourth added detail without opening anything new.

The analysis used interpretative phenomenological analysis, a method built for understanding how people make sense of an experience. I built a separate account for each of the sixteen, which produced 250 experiential statements and 73 personal themes, and only then looked across the cases. Six shared themes emerged, two for each of the study's three questions. I also traced how views moved around each room, and where a participant only agreed with someone else, or took up a word of mine, I did not count that as an independent view. Eight participants, two from each group, read their individual accounts afterwards and confirmed them as fair.

The interpretation is partly mine. I wrote one of the two messages and ran the sessions, so I kept a reflexive journal and logged my own turns in the transcripts to trace where my words reappeared in what participants said. Where a theme depends on something I introduced, I say so.

How the Book Is Arranged

The six themes give the book its first six chapters. The first three follow the order of what a reader does: noticing, weighing and deciding. The next three turn to the conditions around that judgement.

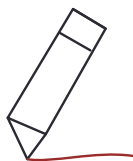
Chapter 1, "The Hand in the Text", shows that participants knew a person by small habits and details only that person could know, and knew the machine by its excess. Chapter 2, "The Ledger of Effort", shows that

readers judged effort by the attention paid to them and their situation far more than by length or time spent. Chapter 3, “Written to Me, or Written at Me”, follows a distinction that ran through the sessions: a reader can accept everything a message says and still withhold trust from the sender, because the use of AI reads as a judgement about how much the reader matters.

Chapter 4, “When Authenticity Matters, and When It Does Not”, shows that the demand for a human hand depended on the stakes, the relationship and the kind of message, and some documents owe the reader nothing personal at all. Chapter 5, “Prosecuting From the Dock”, finds that many participants who criticised AI writing in others also used it themselves, and their standard had shifted from writing a message yourself to standing behind it. Chapter 6, “A Warrant With an Expiry Date”, asks how long any of this can last, since the clues worked only for a reader who knew the sender or had a comparison to hand, and may weaken as the tools learn them.

Chapter 7, “What Writers, Teams and Tool Makers Can Do”, turns the argument into practical advice for writers, for organisations setting rules about AI, and for the people who design writing tools.

The Appendix, “How This Study Was Done”, gives a fuller account of the method and both Sankara messages in full, so that you can reach your own verdict before turning to theirs.



CHAPTER ONE

The Hand in the Text

Readers recognise a person by small habits and personal details, and a machine by too much length, polish and reassurance.

WHEN P16 TRIED TO EXPLAIN HOW A PERSON SHOWS UP IN A piece of writing, they went back to school: “before the era of the internet, we used to have handwritings. Whether in high school or in primary school, we used to know how our colleagues used to write.” Nobody in that classroom had studied anyone else’s handwriting. They simply knew, the way you know a voice on the phone before the caller has given their name.

I have come back to that image more often than to any other in the study. We read the people we work with before we read what they are saying. Most of the time that sense confirms what the sender’s name already told us, so we never notice it working. It becomes visible only when the name at the top says one thing and the prose says another.

Fourteen of the sixteen participants, across all four groups, described some version of this sense. Several prized small errors, many looked for

names, and one read messages aloud to hear whether anything natural survived. The machine, for its part, was recognised by its generosity.

Readers Sense It's AI Before They Can Say Why

Recognition came first, and the explanation, if it came at all, came afterwards. P5 put the moment of doubt before the reading had properly begun:

you just say, no, this is not [colleague's name]. Even if the person has not requested for any help. But maybe just the way the way the message has been constructed, you'll say, I doubt if this one is [colleague's name] before you start your conversation

P5 was certain about the conclusion and vague about the evidence: the doubt is settled, and the reason arrives with a “maybe” and a stumble.

P15 met the same gap from another direction. Their professional English did not give them a word for what they had seen, so they reached into Kiswahili: “the setback is disappointing. Then we have a *kanini* there. a punctuation mark there. It is just AI language.” *Kanini*, a Kiswahili diminutive, did the naming first, and the English followed. The perception had run ahead of the vocabulary available for it.

P7 cited particular words in the machine-written message as proof of its origin, then went back to find them and could not. “There are many words which I saw here, why am I not seeing them now?” they asked. The verdict did not change. The judgement had already been made, and the evidence was being called in afterwards to support it.

People are often right about things they cannot explain, and with both messages in front of them, participants were generally correct about which one the machine had written. It does mean the cues they named, which fill the rest of this chapter, are best read as explanations built after the recognition had happened. They may not be the whole of what produced it.

Larger studies give reason for caution. Jakesch and his colleagues, in experiments with thousands of participants, found that the rules of

thumb people use to tell human from machine writing often fail.¹ Fiedler and Döpke found that when machine-written academic text was polished to a professional standard, fewer than one in five evaluators identified it correctly.² I return to this tension later in the chapter.

Signs That a Person Wrote It

The first surprise was how much participants valued mistakes. Three of them, in separate sessions, treated imperfection as the strongest evidence of a human author available to them.

P5 let the error settle the question: “Because it has some bit of grammar errors, but then it’s like, it’s okay. That one is what the person has written personally from his end or from her end.” In P7’s account the error went further, producing a relation between writer and reader: “I also saw some grammar issues on the first one, which shows that the message came from somebody, it has some sort of connection.” P6 made the trade explicit: “I just told myself I want to relate with this first e-mail despite a bit of grammar issues. I don’t care. I want to have this personal message.”

P6 saw the errors and chose the message anyway, exchanging correctness for presence knowingly. For anyone who corrects other people’s drafts, the tidying we are trained to do can remove what tells a reader that a person was there.

P12 supplied the reverse case, a colleague who writes without greeting or sign-off: “when I get an e-mail from him where he starts with greetings and then he goes to go deeply into explanations, then I can easily tell that this is an AI generated Text.” Here the courtesy is the anomaly. The message is better by any ordinary standard, and that is what gives it away.

Put next to the accounts of error, this shows what participants were reading for: consistency with a person they knew. An improvement was as noticeable a departure as a decline.

The second thing participants looked for was the particular: the detail only someone present could have supplied. The proper name came up in every session. The message I wrote for the study credits one colleague,

Asha, and the team with the pitch; the machine's version names nobody. P1 located the whole difference there:

there is specificities like [...] Asha and the team put together was strong, you know? [...] The second e-mail is a bit general. Yeah, I don't know if, I can't remember reading any name that has been mentioned

P15 agreed: "the introduction of the name Asha, that Asha, you did this, this and this, it looks way personal." The prompt I gave the machine carried only the core facts, without Asha, so the comparison cannot separate the effect of authorship from the effect of the name itself. It does show how much weight readers placed on a single name, and that the machine, working from a brief, had no colleague to name.

P9 found the same kind of evidence at the smallest scale, in a word that had gone missing. My message begins "Just wanted to let you know," without the "I." P9 noticed: "just wanted to let you know. It dropped the I, so there's no I, just just wanted to so you know. AI will never write that." A dropped pronoun is a shortcut people take in a quick note to a team, and a system trained to produce complete, correct sentences does not offer it by default.

P13 told a story that showed the same principle outside the study's two texts. A message had arrived, apparently from a colleague, asking for money: "this colleague was asking me for money. They've never asked me for money [...]. And I knew very well this colleague does not have a spouse." Whoever wrote it could imitate a tone, but did not know the facts of the person they were pretending to be. Such particulars are cheap for anyone who already holds them and expensive for anyone who does not, an asymmetry I return to at the end of the chapter.

Handwriting, Dots and the Sound of a Voice

P16's handwriting image gathers these observations into one idea. Handwriting is formed by habit, stable over time, recognisable to people who never studied it, and very hard to produce on purpose. When P16 carried

the analogy from the pen to the screen, they were placing authenticity in something a writer does not deliberately produce and cannot easily fake.

P10 described being known by such a mark: “Like me, I know people know my message because I use the dots A lot. [...] So everybody will just say this is P10’s message because of the dots.” This account comes from the writer’s side: P10 knows the habit identifies them. It is a small quirk of punctuation, the kind of thing a style guide would correct and an editing tool would smooth away, and to P10’s colleagues it works like a signature.

P8’s example came from speech. The participant described East African colleagues who say a word their own way “shamelessly and with a lot of confidence,” and imagined the alarm that would follow if one of them suddenly spoke differently. Pronunciation, like handwriting, is formed by a particular history, and when it changes without warning, the people who know you notice before they can say why.

P4 had a method for testing whether such a mark was present, and had already been using it on the study’s messages: “if you read it out loud, if you were to read it, that’s actually why I was reading it out loud, just to see, just to bring out the, where is the natural part of it.” A message that is truly someone’s will sound, read aloud, like a person talking.

The dots, the dropped “I”, the way a colleague pronounces a word, the shape of a classmate’s letters: none of these is a decision, and that is what makes each of them worth something to a reader. Machine-generated text, as my participants experienced it, does not carry them.

An African Touch

For four participants, the mark of a person was culturally located, in local idiom and in the ways professional English is spoken and written in Nairobi. This is one of the findings I care most about, and one of the most limited.

P10 named a quality they had not previously had reason to name, and framed its loss as something they would mourn: “there’s that Africanism in terms of... how they use some words and even sometimes they even add

their own local language,” and, a moment later, “I would miss that African touch.” P10 offered this in answer to a direct question from me about African communication, and a moderator’s question can shape the answer it receives.

P8’s account came unprompted and was sharper. Their image of betrayal involved a colleague who “started twanging or speaking like a white person.” The words are blunt, and they place the concern within a longer history of borrowed prestige registers. P15’s *kanini* belongs here too: a Kiswahili word pressed into service to convict an English punctuation mark of being machine-made. The cultural resources a reader brings are part of how they read.

I read these as accounts of four participants’ own workplaces, not as evidence about African communication in general. They do connect to something wider research has begun to document. Agarwal and colleagues found that AI writing suggestions move text towards Western stylistic conventions and reduce the markers of individual and cultural distinctiveness.³ Sourati and colleagues found that when machines rewrite text, most of the meaning survives while the traces from which a reader could infer something about the author are degraded, and that the shift has a direction.⁴

If the cues that tell a reader “a person wrote this” are learned within a culture, as African communication scholars have long argued about voice and presence,⁵ then this homogenising may not fall evenly. It may remove the very cues a Kenyan reader relies on while leaving the text fluent in a register that reader does not hear as anyone in particular. Writing in *The Guardian*, Shariatmadari reported the term “cultural ghosting” for the way these systems correct established non-Western professional English into a standard its writers did not choose.⁶ I make no claim that the effect is worse for African readers than for anyone else; no comparative evidence exists for that. What my participants offer is a first-hand account of what the loss might feel like from inside a Nairobi office, and P10’s word for it was “miss.”

IN PRACTICE: PROTECT YOUR OWN ENGLISH

When you ask a writing tool to polish a draft, read what it changed before you accept it. Research on AI writing suggestions finds that they pull text towards Western conventions and away from the usage that marks where a writer is from. Check that the tool has not “corrected” phrasing that is ordinary and correct in Kenyan or wider African professional English, and that it has kept British spelling (organise, programme, judgement). The participants in this chapter heard their colleagues in these details. A tool that removes them removes the reader’s evidence that you wrote the message.

The Machine Gives Too Much

The machine’s marks, in my participants’ accounts, were large and organised, and most of them came from excess.

P15 saw the scaffolding through the prose:

it looks so structured. It follows a specific structure. To communicate the bad news, show a bit of empathy, motivate the group, explain what are the next steps, close politely

Every step is something a communications trainer might recommend. The problem for P15 was that the steps were visible, and the template underneath showed that nobody had needed to think about this particular team on this particular afternoon.

P4 registered the same excess as a surface placed in front of something, repeating a phrase as they spoke: “there’s no shortening of phrases. Like it’s picture perfect. It’s picture perfect. It just creates a [...] facade.” A phrase a person would shorten when writing to colleagues had been left at full length, and the completeness read as something put up for show.

P14 found comfort turning into alarm, describing the machine’s message as “too reassuring to an extent that it’s almost trying to, you know, make the team feel like it’s the end of the world.” The consolation inflated the loss it was consoling.

P5 identified the pattern at the level of the sentence. The machine, they said, “will mention for you things, a number of them, two, three at once,” where a person “will just mention one thing as you move on.” The machine’s version of my message thanks the team for their “hard work, commitment, and professionalism,” and praises their “talent, creativity, and dedication.”

In each of these accounts, generosity reads as failure. The machine supplied more warmth, more structure and more reassurance than the occasion required, and the surplus gave it away. My message ran to 147 words; the machine’s, written from a prompt with the same core facts, ran to 222. Research on machine-written text describes the same tendencies in general terms: more formulaic language, more consistent grammar, less variety,⁷ and an emotional tone pulled towards the positive and agreeable at the expense of doubt, frustration and the other registers people actually feel.⁸

The machine’s version of my message contains a dash in its third paragraph, produced by the tool and not added by me. In two sessions, participants treated a punctuation mark of this kind as evidence of machine authorship, and P15’s *kanini* was one of them. Yet the dash, like the list of three and the polished cliché, is an old feature of human prose. Hardaker, quoted by Shariatmadari, observes that most people identify machine-written text correctly only around sixty per cent of the time, and describes the cues they rely on as simplistic rubrics applied indiscriminately.⁹ A reader who convicts a message on a dash may be right about that message and wrong about the next one.

Jakesch and his colleagues found that people associated first-person pronouns, informal contractions and references to family with human authorship, even though machine-written text contained these features just as often.¹⁰ My participants reached for cues of this kind, and on that evidence their rules of thumb should have failed. With both texts in front of them, they mostly did not. Three reasons are possible, and the data do not let me choose between them: a comparison made the difference visible in a

way an ordinary inbox would not, as one participant pointed out; the machine's message was a single system's default response to a brief instruction, the most patterned form such text takes; and participants were reading a kind of message they write every day. Readers attending closely, with a comparison available, in a familiar genre, told the two apart consistently, and explained their judgements with cues that other research has found unreliable.

FIGURE 2 The message I wrote, as participants read it

Good afternoon team,

I hope you are well. Just wanted to let you know ^① that we have just heard from Sankara on our bid (reapplication) that they will be proceeding with another agency. I know this is a blow to many of us in the team who put in a good effort on this account, and definitely it would have been nice if we kept it. The pitch Asha and the team ^② put together was strong, and from what I understand so far it came down to price. I have requested the full feedback so that we can see what to improve on for next time, and I'll share it once it comes.

I propose we meet on Monday at 9 to regroup and talk through what each of us picks up next.

Have a restful weekend, and we'll take it from there on Monday.

Best,
Siddondo

- ^① "It dropped the I, so there's no I, just just wanted to so you know. AI will never write that" (P9)
- ^② "the introduction of the name Asha, that Asha, you did this, this and this, it looks way personal" (P15)

FIGURE 3 The message ChatGPT wrote, as participants read it

Dear Team,

I am writing to inform you that we have received confirmation this afternoon that the Sankara account has been awarded to another agency. Unfortunately, despite the strong pitch prepared by our team, the renewal decision was based primarily on price.

I understand that this news may be disappointing. I would like to thank each of you for the **hard work, commitment, and professionalism** ❶ you demonstrated throughout the pitching process. Although we were unsuccessful in retaining the account, the quality of our proposal reflected the high standards of this agency.

While we cannot change the outcome, we can control how we respond. **This setback is disappointing—but** ❷ it does not diminish the talent, creativity, and dedication that each of you brings to the agency. **I encourage everyone to remain positive and focused.** ❸ Although losing an important client is challenging, it also presents an opportunity for us to reflect, learn, and continue delivering excellent work for our existing and future clients.

To discuss the way forward, we will hold a team meeting on Monday at 9:00 a.m. During the meeting, we will regroup, review our priorities, and reassign work where necessary to ensure a smooth transition.

Thank you for your continued dedication and cooperation. I appreciate your efforts and look forward to working together as we move forward.

Kind regards,
Siddondo P. A.

- ❶ “will mention for you things, a number of them, two, three at once” (P5)
- ❷ “the setback is disappointing. Then we have a *kanini* there. a punctuation mark there. It is just AI language” (P15)
- ❸ “too reassuring to an extent that it’s almost trying to, you know, make the team feel like it’s the end of the world” (P14)
- On the whole: “it looks so structured. It follows a specific structure” (P15); “It’s picture perfect” (P4)

IN PRACTICE: COUNTING THE EXTRAS

The participants caught the machine by its surplus. In the drafts I edit, most of that surplus comes from a handful of habits. Items are grouped in threes for rhythm, as in “hard work, commitment, and professionalism”. Opposed adjectives are paired for effect: “small but meaningful”, “simple yet profound”. A short line is added at the end of a paragraph to underline a point the paragraph has already made. Mark each one in your draft. Keep a list of three only when every item is true and specific, and allow yourself one per page. Replace the adjective pairs with what actually happened, and fold the underlining line into the sentence before it or cut it.

Why the Small Things Carry Weight

Two accounts run against the picture I have drawn.

P13’s trust rests on the institution. For this participant, a message is genuine when it follows shared conventions: a house font everyone uses, a copy list that is respected, “patterns that have been cultured over time amongst colleagues.” Conformity, for P13, is the mark of the real thing, and a stranger’s email, belonging to no group they know, offers them nothing to read.

P12 inverted the tempo the rest of the chapter assumes. Most participants saw the person as the one who takes time and care. P12 saw it the other way round: “as humans, we like to get straight to the point. finish things fast.” For P12, the machine’s tell is that it lingers, and a person is recognised by their impatience.

Both readings still check a message against a known pattern: an institution’s in one case, a busy person’s pace in the other.

Walther and Parks observed that when people present themselves through a screen, they have unusual control over the impression they give, so a receiver learns to trust information about a person in proportion to how hard it would be for that person to fake.¹¹ They called this warranting. A claim such as “I care deeply about this team” costs me nothing to write and carries little weight. Something that would take real effort or

special access to produce carries more. DeAndrea later broke this difficulty into two parts: the effort it would take to fabricate a piece of information, and the access a person would need to produce it at all.¹²

My participants applied this reasoning without using the word. The cues they trusted required access they could check: a colleague's name in a message about that colleague's work, a fact about someone's family, a shared history. The fraud P13 described collapsed on a detail of domestic life the sender could not have known. The cues they distrusted were the ones anyone can produce, a machine included: fluency, structure and courtesy.

The study's proposal is that a writer's idiosyncrasies belong with the trusted cues. The dropped "I", the dots, the particular grammar slips, the habit of saying one thing and moving on, are high in both effort and access, because they come from a particular person's history with language. Participants themselves noticed that a system can be instructed to add errors, so I do not claim these marks are impossible to fake. I claim they are expensive to fake well, and cheap for the one person who already has them. When a text is fluent, correct and stylistically unremarkable, the reader has nothing that ties it to a particular life. I call that condition warranting failure: the cues that link a text to a person are simply absent.

This explains why the machine-written message could be judged competent and still be rejected. Coman and Cardon, surveying close to nine hundred working professionals, found that AI-generated workplace messages were rated well on every functional measure, while perceptions of sincerity and caring fell when AI authorship was disclosed.¹³ My participants described the same split from the inside: they mostly accepted the machine's quality, but were looking for evidence that a particular person had made it, and could not find any.

These are sixteen people in one city, reading one kind of message, with a comparison to hand, and their accounts describe experience, not how often anything happens. The warrant they describe also depends on a reader equipped to read it. A colleague who knows your dots will notice when

they disappear; a stranger has no baseline. That condition, and the possibility that these cues will stop working as the tools learn them, is the subject of a later chapter.

What This Means for the Writer

The first lesson is to keep the particulars. The features that convinced my participants were the ones a machine could not supply from a brief: a colleague's name attached to what that colleague did, a reference to something the team shares, an honest account of what is not yet known. If you draft with assistance, put these back by hand, since a tool is least likely to offer them. A message that names Asha tells the team that the writer knows who did the work, and no amount of thanks addressed to "each of you" can say the same.

The second is to be slower to correct your own habits than your training may suggest. Participants valued small errors, characteristic punctuation and the shortcuts people take when writing to colleagues, because these were the marks of a person they knew. In many documents accuracy matters most. But when a message is meant to come from you, especially when it carries news that affects people, the smoothing that editing tools perform can remove what tells your readers you were there. If your colleagues know your message by its dots, think twice before letting a tool take them out.

The third is to watch for excess. Participants recognised the machine by what it added: the visible template, the extra paragraph of reassurance, the list of three virtues where one would do. Before sending anything drafted with help, read it aloud as P4 did, and listen for the places where no person would talk like that. Cut the sentences that explain what the reader already understands, and match the length to the occasion. In the study, the shorter message was the one readers believed.

The fourth is to remember that consistency is what your readers are checking. In P12's account, a sudden politeness betrayed the machine. If a tool makes your messages sound unlike the ones you usually send, the

people who know you will notice the change, even if they cannot say what it is. Their judgement will arrive before their reasons, as it did for P5 and P7, and once formed it is unlikely to move.

Finally, your voice carries more than you may realise, including the local idiom of the place you write from. Some of my participants feared losing the African touch in their colleagues' messages and recognised a borrowed register as a kind of departure. If you write professional English in Nairobi, the ways it differs from a standard model are worth keeping where they come naturally. For the colleagues who share them, they are part of how you are recognised.



CHAPTER TWO

The Ledger of Effort

Readers repay a writer's effort with their attention, and withhold it when the effort seems missing.

PARTWAY THROUGH THE SECOND OF THE FOUR DISCUSSIONS, P6 turned the conversation back on the group. The four participants had read both versions of the message about the lost Sankara account and agreed on which of the two had cost its sender anything. P6 wanted to know what that agreement rested on:

one is long, but we've unanimously agreed that there was no effort [...] The other one was short, But all of us can agree that so much has gone into it. What is that so much? (P6)

P6 pressed the point later, asking "what do we mean those invisible efforts? that go into a piece of communication that end up being of priority or of preference or of importance to the recipient". It was a fair question, because the two texts on the screen seemed to argue the other way. The message I had written was 147 words long, in three short paragraphs. It said the account had gone to another agency on price, that the pitch Asha and the team had prepared was strong, that I had asked for the full feed-

back and would share it, and that we would meet on Monday at 9. The version ChatGPT produced from the same facts ran to 222 words in five paragraphs. It thanked the team for their “hard work, commitment, and professionalism”, encouraged everyone to remain positive and focused, and called the same meeting. By any count you could run on the page, the second message contained more.

P6 answered the question with an arithmetic that runs backwards: “If you write one statement [...] I’ll be more connected to it. And I’ll take it as if you had so much effort. Then you write 3 pages that I feel like this is not you.” The more a message contains, the less the reader believes went into it.

Eight of the sixteen participants spoke about effort, and their accounts form a single picture. They judged effort by the attention a message showed to them and to the situation, and not by length or by the time they imagined the sender spending at a keyboard. They kept that judgment apart from whether the sender meant what they said. And some described adjusting their own effort to match. I call this the ledger of effort. The image is mine; the accounting belongs to the participants.

Effort in the Thinking

P16 answered P6’s question by splitting the word in two:

Of course, effort in terms of the energy expended in the drafting or in the thinking (P16)

Drafting is the work you do on a text: choosing words, building paragraphs, checking the grammar. Thinking is the work you do on the situation: what has happened, who will read this, and what they need. A machine can now do almost all of the first. The question is whether anyone has done the second.

P16’s group was the one that saw the machine-generated message first, so for P16 the second email was mine. P16 reached a conclusion that runs against the obvious reading:

in the second e-mail, brief as it may be, there is a greater thinking through [...] To me, there's more effort in e-mail 2 in terms of the desired impact of the messaging compared to e-mail one (P16)

What P16 found in the shorter text was evidence that someone had considered what the message would do to the people receiving it.

P16 could make this judgement because they had already seen through the longer message to the act that produced it. Before I had said anything about how either text was made, P16 reconstructed aloud the instruction they believed had generated the machine version: "He went probably to ChatGPT and told them, you know, I work for company XYZ. We have received information from Sankara that a bid was not successful. Draft an e-mail to the team to convey this outcome." That guess is close in kind to the short prompt I had in fact given the tool. Once a reader pictures the sender typing a few lines into a box, the length of what came back stops counting as the sender's work. What remains is the thinking done before the box was opened, and in P16's reading there was very little. A long message can be produced without any attention to the loss and what the team would need to hear, and a short one can be full of it.

In my own work as a communications specialist I have often seen the opposite assumption: that a longer, fuller message shows more respect for the reader. When a draft took an afternoon, that was reasonable, because three pages meant an afternoon spent on you. The participants had already stopped making that inference. They knew that pages now cost nothing, and they looked for the things that still do.

Feeling Before Writing

P10 moved effort earlier still. In one session I offered a sentence of my own comparing the effort behind the two texts, which I should not have done, and I do not treat it as evidence. What matters is that P10 declined the word I had used:

Not really effort. It's not really, okay, I don't know if you want to call it effort (P10)

and then worked towards what they meant in its place: "the person actually sat down and felt the disappointment and the pain and also thought through how he's going to respond to his team and also manage the emotions around the team." The sender has two jobs before writing a word: to absorb the loss, and to think about how the team will absorb it. Neither leaves a mark on the word count. On P10's account the obligation is met, or not, before the writing starts.

P4 read my message the same way, using their own capacity as the measure. The text, they said, "makes me feel like even the manager themselves Is. disappointed and really cannot say much at this point". And then: "if I was to respond, I would also not be able to write so much because I'm also disappointed." A person who has just lost the agency's largest client does not usually have five paragraphs of encouragement ready by the afternoon. The short message matched what P4 imagined a disappointed person could manage, and so it read as written by one.

The novelist Jeanette Winterson, quoted in a recent piece in *The Guardian*, puts the same condition in physical terms. She cautions that machines "do not share our reality, not least because they don't have a limbic system. Humans cannot have a thought without a feeling".¹ P10's requirement is the professional version: before a leader writes about a disappointment, they are expected to have sat with it.

P8, speaking about messages from colleagues, described the failure case. What a sender hands to a tool is their feeling: "I can always tell that this person has decided to outsource their feelings to through technology or to an AI platform. And this is not really them." P8 also named what the delegation costs the sender: "You are not learning anything because you did not write it in the 1st place. So yeah, that for me, in that regard, I find it unprofessional and lazy." That last word turns up in the published research as well.

P16 found an image for what arrives when the feeling and the thinking have been delegated: “they respond with echoes and echoes of explanations and details. That’s probably even necessary, but not very useful.” P16 allows that the explanations may be correct; the trouble is that they answer nobody in particular. P16 traced the fault to a “lack of that personal drive really to, first of all, apply yourself to what you’ve been asked about”, which leaves the reader with “a sense of not only despair, but also of frustration”.

These accounts separate two questions that are usually asked as one. The first is whether the sender means what the message says, which is a question about sincerity. The second is whether the sender has looked closely at this situation and these people, which is a question about attention. The machine-generated message is full of statements of feeling. It says “I understand that this news may be disappointing”, thanks each member of the team, and assures them the setback “does not diminish the talent, creativity, and dedication that each of you brings to the agency”. None of the participants in this theme took those statements as evidence of anything. P4 found evidence of feeling in the shortness of my message, and P16 found evidence of thought in its “desired impact”.

On my reading, then, a sender can be entirely sincere and still fail the test these readers apply. P8’s phrase allows for that: a person who outsources their feelings may still have them. What the reader cannot find, in a message assembled from a brief instruction, is any trace of attention to this loss and these people.

Less Effort In, Less Goodwill Back

P15 turned the judgement into an account in the literal sense:

surely we are paying, this person is being paid because for me as an accountant, like all this salary we are paying for somebody to just generate something, a random response via ChatGPT and send to us (P15)

The example concerned a supervisor, and P15 describes the supervisor's work as paid for and not delivered. This is the ledger at its plainest: tracking what a message cost its sender, setting that against what was owed, and noticing the shortfall. P6 asked what the invisible efforts were, P16 priced them in thinking, P10 in feeling, and P15 in salary.

The reckoning also changed what readers did. P15 answered withheld effort with withheld attention: "I didn't even take time to look at it. I just ignored it." Applied to the two texts, a generic message would receive a generic answer: "responding a generic way. Thank you, this is well noted". For the message I had written, "you can put more feelings there, because it feels like a conversation".

P7, in the second session, carried the accounting into work not yet done:

I may not want to put in a lot of effort in this in the other pitch, because after all, my boss also does not put effort in even giving us feedback (P7)

Of my message, P7 said: "the first one, there was a personalized effort. So good effort went in." Of the other: "The second one, I think the boss was not serious." Its sender, P7 said, "was just beginning to check a box and move on". The next pitch is work the message never mentions, and P7 was proposing to carry the deficit into it.

P4 described the withdrawal happening inside the act of reading. After I had told the group how each message was produced, I asked whether that changed anything:

I will probably not read them with so much detail because I know it's not, you know, there's nothing genuine about it. So I would just want to get out the action items (P4)

The action items serve the reader's own needs. The acknowledgement and thanks would have carried the relationship, and P4 proposed to stop reading them.

P3, in the first session, reasoned from the sender's investment to the significance of the matter:

If this bid was that deep, then you would take your time to send an e-mail to me that is from you, from your brain. Now, if you generate AI, it means even me, I shouldn't be that worried about this bid because it really is not that deep to even you as a team lead (P3)

P3 takes the sender's effort as information about how much the event matters. A manager who reaches for a tool to save time on a hard message may find that the team reads the shortcut as a statement about the stakes.

I take these four accounts to be the principal finding of the theme, and of the study. Perceived effort set these readers' own spending: on reading the message, on replying to it, and in P7's case on later work. In the dissertation I called this readers enforcing a communicative effort contract, the expectation that a sender invests effort in what they send. The phrase is mine; the conduct was described by the participants.

These are four people in one city, discussing two texts and reporting what they had done or said they would do. None of it is an observed effect on anyone's work, and P7 said nothing further after I disclosed how the messages were made, so I do not know whether the disclosure changed P7's view. What the accounts offer is a proposed mechanism, seen from the inside. Whether it happens widely, and at what cost, is a question for studies built to measure it.

Even so, most discussion of AI writing tools at work counts the time they save the sender. If these accounts describe something real, part of that saving is paid back at the other end, in attention the reader withholds.

IN PRACTICE: OPEN ON THE FACT

Readers looked for signs that the writer had paid attention to them and to the situation. Sentences that announce what is coming spend the reader's attention without giving anything back. The ChatGPT message in this study spends its first fourteen words telling the team that news is on its way ("I am writing to inform you that we have received confirmation this afternoon that...") before it reaches the news. Delete announcement sentences and open each paragraph on a fact, a name, a number or an action. It is usually the quickest single improvement to a draft, and it rarely loses any content.

Where Readers Disagreed

P16 found more effort in the shorter message, so effort cannot simply be read off the page. P16 reached that view after reconstructing how the longer message was made, so the judgement rested on a belief about the act behind the text, and length counted only through that belief. A reader who had not formed that belief might have weighed the two messages differently, and the study cannot say how.

P10's refusal of the word "effort" pulls the other way. The whole theme is built on that word, and P10 told me it was the wrong one. What P10 asked of a sender was emotional work undergone in advance, which is not work performed on a text, and I have tried to report it in P10's terms. Between them, P16 and P10 locate effort in thinking and feeling that happen before the first word is written. What reaches the reader is whatever trace of them the text carries.

What Other Studies Have Found

The philosopher Paul Grice argued that people in conversation assume each other to be cooperating, and that when a speaker is conspicuously brief, or conspicuously elaborate, the listener infers something from the choice.² A carefully chosen word where a generic one would do signals in-

vestment; a message indistinguishable from any other signals that this exchange was not thought worth individual attention.

Dan Sperber and Deirdre Wilson sharpened that idea into a single principle. They defined relevance as a ratio: what a reader gets out of a message, set against the effort it costs to process. When someone addresses you, they make an implicit promise that the effort of attending will be repaid.³ This is the most useful way I know to see what the participants were describing. Machine generation lowers what a message costs to produce almost to nothing, and the text it produces tends to be longer and more explanatory than the occasion needs, which raises what it costs to read. The reader is asked to spend more to receive less, from a sender who, as far as the reader can tell, spent very little. The withdrawal the participants described restores the balance by the only means a reader has, which is to reduce their own side of it.

Researchers who study language models have made the related point that these systems cannot invest effort in the relational sense at all, since they produce text without any stake in the people who will read it.⁴ Whatever effort a reader finds in such a message has to come from the person who used the tool.

In an online experiment on AI-mediated email, Liu and colleagues found that trust in a writer fell when AI involvement was indicated, and that participants explained the drop by reading AI assistance as a sign of less effort and a less authentic message. They were least willing to accept AI help for sensitive messages such as condolences.⁵ Zhu and Molnar, in two experiments with 1,301 adults in the United States, found that senders known to have used AI were rated as lazier, less sincere and less effortful than those who wrote their own messages.⁶ P8's word, "lazy", is the same word. Zhu and Molnar also found that when AI use was not disclosed, readers rated the messages as highly as human-written ones and rarely suspected anything.

In workplace writing, Coman and Cardon report a similar split. AI-assisted messages kept their ratings for professionalism, while ratings for sin-

cerity and caring fell once AI authorship was known.⁷ A pre-registered study by Sahebi, Formosa and Bankins, which included workplace email among its contexts, found an AI penalty on trustworthiness, authenticity, and what they call knowledge uptake, the reader's willingness to use what the message contains.⁸ That third measure is the one closest to this chapter, because willingness to use the content is what P4 and P15 described withdrawing.

These studies agree with what the participants said, but each measures the effort judgement after AI involvement has been disclosed or labelled. They do not show how readers infer effort from the text itself when no one has told them anything.

The participants add two things. Several made the effort judgement before any disclosure: P16 reconstructed the prompt, and P7 read one message as personal effort and the other as box-ticking, before I had said how either was made. They also described what they did next. The published studies record evaluations and, in Sahebi and colleagues' case, a drop in uptake. The participants here supplied the reasoning that might produce such a drop: a judgement about what the message cost its sender, followed by a decision to match it. That step, from what audiences think about AI-mediated messages to what they stop doing when they receive one, is the part of this study I most want others to test.

The novelist Gary Shteyngart, quoted in the same *Guardian* piece, describes the bargain from the writer's side: "a kind of implicit bargain between writer and reader where you know the work that you're getting is generated by a human being".⁹ He was speaking of literature; the participants in this study applied the same bargain to an email about a lost client.

What This Means for the Writer

The first lesson is about where to spend your time. The work that counts happens before you write. Before you open a draft, or a tool, settle what has happened, what it means to the people you are writing to, and what

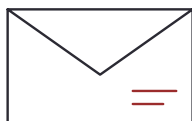
they will need from you when they finish reading. That is the thinking P16 found in the shorter message and the feeling P10 asked a leader to sit with. A tool can help with the drafting afterwards, but readers who know these tools will look for evidence that you did this part yourself.

The second is about length. Do not take length as a sign of care, and do not expect your readers to. Participants read surplus explanation, reassurance and structure as a failure of attention. Match the length and form of a message to the occasion. If a draft has come back from a tool longer than you would have written it, cut it to the length you would have written.

The third is about what you put in. Statements of feeling did not persuade these readers; signs of attention did. The features that show attention are the ones a tool cannot supply from a brief instruction: the name of the colleague whose work is in question, the thing you have already done about the problem, the plain admission that you do not yet know the full story. Where you have drafted with help, put these back by hand. And own what you send. Several participants suggested checking whether the sender can remember what the message said and answer a question about it. Apply that test before you press send.

The fourth is about what your reader does next. A generic message received a generic reply, a message judged machine-made was read for its action items, and one reader said they would put less into the next piece of work. If you are a manager, the time a tool saves you may be repaid in your team's attention. Bad news, condolence, feedback on someone's work, and thanks for effort were the messages these participants were least willing to see delegated. Where the purpose of a message is relational, write it yourself. Where it is only informational, a tool can serve.

You can keep using the tools; most of the participants used them themselves. What these accounts ask is that you be clear about which part of the work your reader is counting. On the evidence of these sixteen readers, they are counting the attention you paid to them, and they keep their own accounts accordingly.



CHAPTER THREE

Written to Me, or Written at Me

Readers can believe a message and still feel that no one wrote it to them.

ONE PARTICIPANT, P6, TOLD THE GROUP ABOUT A PIECE OF feedback from a supervisor. It was long, longer than the work required, and something about it did not sit right. The participant looked into it. “The feedback was quite lengthy beyond what we need. And interrogating further, I realized it was actually a copy of ChatGPT.” What came next was a question, and the participant was careful to tell us that it was never spoken aloud:

so why am I even reporting to you? Of course, I did not ask him that question, but I just had a reflection (P6)

The feedback itself may well have been sound. Nothing in the account suggests the participant found the content wrong or ignored it. The damage was to the relationship. In the participant’s words, the supervisor had “instantly lost their authority. that they can’t really dictate what I need to

do in that work [...] everyone is sort of playing on the same level ground.” A reporting line had gone flat in one reader’s mind, and the supervisor almost certainly never knew.

A message can do its job as information and still fail as an act between two people. The reader keeps the content and revises the sender, and because the reader has no reason to announce the revision, the sender carries on as if nothing has happened. Eleven participants contributed to this theme. Two of them read the same material quite differently, and their readings are here too.

Believing the News

The message every group read announced bad news: the agency had lost a large client account, the decision had come down to price, and there would be a meeting on Monday morning to regroup. Participants called the lost account “the bid”, and nobody in any session doubted that the bid had been lost.

When the group turned to trust, P14 began, stopped and began again: “so one I will, I will, I will trust, I will trust the second one, but all the OK trust, I don’t think trust for me is is is is any different.” Over the next few turns the participant kept at it. First they moved the question away from the message altogether: “it doesn’t tell me much about whether I should trust it or not. It just tells me more about the person conveying the information.” Then they landed on it:

So one, I may trust all of them, but I will feel more as part of the team with a second message (P14, referring to the human written version)

I regard this as the plainest statement of the theme anywhere in the data. The participant’s belief in the content is untouched. What moves is whether the reader feels inside the group the message speaks to. That sentence separates two things most of us treat as one when we talk about trust at work: whether I believe you, and whether I feel you were talking to me.

P16 put the question that organises this chapter, and did it by first setting aside everything a style guide would check:

the English is okay, everything is perfect, the reporting is fantastic. But how does it make me feel if I'm a team member? Is it written [...] to the team or written for the team (P16)

The participant then found the failure in the grammar. The machine-generated message, they said, "is almost written in. in a third person aspect", and "the writer is so far removed from the message they're conveying, but it's very formal." Read the generated text with that in mind and you can see what P16 meant. It thanks the team, praises them and encourages them, but it speaks of them from a slight distance, as a body to be informed and steadied, in the register of an announcement. A message can be correct, complete and entirely about its readers and still never turn to face them.

P5 described the same thing from the other side. Reading the human-written message, the participant felt it was addressed "to the people, rather than employees", and felt that address pulling a response out of them:

The fact that the boss has taken it personal and has written to us with that feeling, I will also want to be in the same context and also respond like that with that respect (P5)

P5 moves straight from the message to the reply, and would answer in kind. Being addressed as a person creates an obligation to respond as one. If P5 is right about themselves, a team that feels written to will reply with something of the same regard, and a team that feels written at will reply as employees. The content is the same in both cases, but the working relationship it travels through is not.

Feeling Managed

For several participants, the machine's message did more than keep them at a distance. They read it as an attempt to manage them, and they resented it.

P8 gave the fullest account, starting with what the human-written message had offered:

they were straight with me. They trusted that I wear big boy pants and I can take it, but also that they have acknowledged, the team leader has acknowledged the work that we have taken (P8)

P8 called this “humanized leadership in how that e-mail is crafted. The leader is not trying to make it sound like they’re superhuman.” Then the participant turned to the generated version, with its paragraphs of thanks, its reassurance that the setback did not diminish anyone’s talent, its encouragement to remain positive and focused. P8 read all of that as an assessment of their capacity:

the leader doesn’t think I am adult and professional enough to take bad news. So the scaffolding I think that you are talking about is a bit too much. And, and, and it it it insults my intelligence as someone who had made significant contribution to that pitch (P8)

P8 spoke fluently for most of the session. The stumble comes exactly where the offence is being named, and I read it as a sign of how much the point mattered.

P4 reached the same place more gently: “I feel a bit more emotionally managed, basically.” P3 put it down to a misreading of the moment, saying the message “doesn’t feel like it’s coming from a team leader who is reading the room properly.” In each case the trouble was an excess of care that read as a lack of respect.

I did not expect this when I designed the study. In my own work as a communications specialist I have often been asked to add warmth to a difficult message, and the usual instinct is to add more thanks, more reassur-

ance, more framing of the setback as an opportunity. The generated message did all of this competently, and many organisations would approve it without a second look. For these readers it backfired. Reassurance the reader has not asked for, at a length the occasion does not need, tells the reader that the sender sees them as fragile and in need of handling.

The two versions were not identical in content. Mine named a colleague, Asha, whose pitch it called strong, and it promised to share the client's full feedback when it came. The generated version did neither. Some of what P8 heard as acknowledgement may come from a name and a promise the other text simply did not contain. The study cannot separate those differences from the manner of writing, as I explain in the methods account at the end of the book. What the participants describe is still real as experience: a reader who feels named and kept informed in one message, and handled in the other, will judge the senders differently.

IN PRACTICE: WATCH THE LAST PARAGRAPH

Much of the feeling of being “managed” sits at the end of a message: the call to stay positive, the thanks for “continued dedication and cooperation”, the line that sounds as if it came from a corporate values statement. Endings written to sound institutional, or to be quoted, read as written at the reader. End on the last real point instead, such as the meeting time, the next step, or the thing you will do and by when. If you need to ask something of the reader, make one specific ask, once.

Imagining Who Sent It

When a message fails to address its reader, the reader does not leave the gap empty. Participants built a sender to explain the message, and the figure they built was rarely flattering.

P6 stated the mechanism earlier in the same session that produced the supervisor story: “It might communicate the message, but it starts to lose that meaning which was intended because I attach meaning to you as a person.” The message survives. The meaning it was sent to carry does not,

because the person it was attached to has gone missing from the text, and the reader fills the space with an account of why.

For P7, the explanation was incapacity: “Of course, I lose confidence in them because they are not able to communicate really. As them.” The participant then took it from competence to character: “I view my boss to be someone who is not serious.” Here, unlike for P14, the damage reaches the sender as a person and not only as a writer.

After the two texts had been discussed, I told each group which version I had written and which ChatGPT had produced. P1 said the disclosure changed nothing for them. Both versions were signed by the same sender, Siddondo, yet P1 gave the author of the machine-generated email a name of their own, as if a second person had written it:

I strongly believed that the e-mail 2 [...] Siddondo Siddondo 2. was insincere, was lazy, and lacked authenticity (P1)

As I read it, the participant had no quarrel with the Siddondo whose message they took as human, so they split the sender in two and gave the verdict to the copy. Insincere, lazy, lacking authenticity: each is a judgement about a person. The reader needed someone to hold responsible, and supplied one.

What P8 and P4 described as an injury to their regard for the sender, P6 described, in the supervisor story, as the collapse of a reporting relationship. The supervisor’s authority depended, in the participant’s mind, on the supervisor being the one who had looked at the work and judged it. Once a machine seemed to have done the looking, there was no reason to defer, and because the question stayed unasked, nothing was raised that could have been repaired.

Most of what participants described here happened silently. P8, P7 and P6 did not, on their own accounts, say anything to the sender. A reader who withdraws regard has no occasion to report it, so the harm that matters most here may be the harm that never surfaces.

Learning which message the machine had written affected participants differently. P1's verdict stayed the same. P2, whose reading I come to below, found that it cost the message nothing. Several participants changed the terms of their objection after the disclosure more than its strength. P8, for example, having called AI-mediated writing inauthentic and lazy, afterwards allowed that polishing the reader could not see would give no difficulty. That shift belongs mainly to a later chapter, but it suggests that what offended P8 was being able to see the machine, and so to see that the sender had not faced them. P7 said nothing after the disclosure, so I make no claim about whether it moved them.

Two Readers Who Saw It Differently

Not everyone experienced the generated message as a failure of address, and two dissenting readings are as much part of the evidence as anyone else's.

P2 found the AI-generated message the more caring of the two. For P2, the task of a message like this was to keep the team steady after bad news, more than to speak to each reader as an individual. "The second one I feel has more of a humane aspects to it, not puncturing the confidence of the team," the participant said, and later, "I'm looking at grief, grief management." Where P8 heard padding, P2 heard protection, and the shorter human-written version risked sounding hasty.

Twice P2 slipped into the role of the person who had written the two messages, once speaking of "maintaining my team", and once describing how the texts were made as though they had made them: "Then I used a bit of AI to do the second one to bring a bit of the human feel." From the sender's chair, the generated text looks like a leader doing the responsible thing for a shaken team. From the reader's chair, several participants felt it as being managed. For P2 the disclosure cost nothing, since a human intention had been present throughout and the tool had simply helped carry it out.

P10 found the generated message hostile. Asked what they felt from it, the participant turned down the obvious word three times before settling on a stronger one: “The person is pissed. In this second e-mail, the person is really pissed.” The participant later moderated this to a flatter description, and I keep both positions on the record. A formal, distant message about bad news does not announce its sender’s state of mind, so the reader supplies one. P1 supplied laziness, P7 a lack of seriousness, and P10, at first, anger. P2 shows the alternative: a reader who assumes a caring human intention behind the text, and then finds care in it.

These two accounts mark the limits of the chapter. The pattern is strong in this data, but it comes from sixteen professionals in Nairobi reading one kind of message with a comparison in front of them. A reader who takes the sender’s view, or who reads a long consoling message as the right response to a loss, may see warmth where others see management. Writers will meet both kinds of reader, often in the same team.

What Other Studies Have Found

Hancock and colleagues gave this area its working name, AI-mediated communication, when the tools first reached professional use.¹ Since then a steady run of studies has found that trust in the sender tends to fall when readers believe AI was involved. This study adds a view from inside that fall.

Most organisational researchers still work from the model of Mayer, Davis and Schoorman, who described trust as a willingness to be vulnerable to another person, resting on three judgements: whether they are able, whether they mean well towards you, and whether they hold to principles you accept.² Coman and Cardon, surveying 887 working professionals, found that AI-generated messages were rated professional, effective and competent, but that when AI authorship was disclosed, perceptions of the sender’s sincerity and caring dropped.³ In Mayer’s terms, ability held up while goodwill and integrity did not, and participants here described the same pattern. P16 called the English okay and the reporting

fantastic, while P1 called the sender insincere and lazy, and P7 called the boss not serious.

Cardon and Coman also found that professionals saw their supervisors as sincere far less often when those supervisors leaned heavily on AI, even among respondents who could not reliably tell which messages were AI-produced.⁴ Schilke and Reimann, across thirteen experiments, found that people who disclosed their AI use were trusted less, because they seemed less legitimate, and that the effect was stronger when someone else exposed the use.⁵ P6's supervisor was exposed by the participant's own investigation, and the authority went with it. Sahebi, Formosa and Bankins found the same penalty in workplace email on trustworthiness, authenticity and knowledge uptake, the reader's willingness to use what a message tells them.⁶ That is close to P5's point from the other direction: a reader who does not feel addressed gives back less. Dorigoni and Giardino found that the same supportive message was received differently depending on whether readers believed a person or a system had written it, so warmth depends on who the reader thinks is behind the words.⁷

Logg and colleagues documented algorithm appreciation, a tendency to trust algorithmic judgement over human judgement in some tasks, though it is weakest where a task calls for character and ethical judgement.⁸ A team leader breaking bad news sits in that second group. Coman and Cardon also note that readers under time pressure may prefer a clean, uniform voice that is quick to process.⁹ P2's reading suggests a version of this.

Where this study adds something is in what P14 described. The trust literature asks readers to judge the sender. P14 was judging their own position, whether they stood inside the group the message addressed. I have called this relational membership. It can be lost while every judgement of the sender stays intact, and surveys that ask how sincere or caring a sender seems are not built to catch it. Dirks and Ferrin's review found that trust in leaders depends on the quality of the leader and follower relationship as well as on visible competence.¹⁰ If that relationship is built message by message out of the experience of being addressed, a message that fails to

address its readers draws it down. P14 felt less part of the team over a single email.

Credibility researchers separate the credibility of a message from the credibility of its source.¹¹ The study held the facts and the named sender constant across both texts, which is why the separation P14 describes is visible at all. In real life, a reader who stops trusting the sender may eventually doubt the content too, and this study cannot say when. And Nkembuh found that in remote teams AI mediation first got in the way of trust but led to stronger working relationships where it was handled openly.¹² The readers here had bad news from a team leader with no explanation of how it was made, which is close to the least forgiving set of conditions there is.

Walther and Parks, who first described the warranting principle, held that we give most weight to cues that are hard for the other person to fake.¹³ Earlier chapters argued that a writer's small habits and particular knowledge work as that kind of cue. Those cues also show the reader that the sender was there, and was looking at this team and this loss when they wrote. When the cues are missing, readers keep believing the facts and stop believing they were the ones being spoken to.

My own position shaped how I read this material. I came to the study as a practitioner who spends much of his working life writing for leaders, ready to sympathise with a busy manager reaching for a tool to soften a hard message. The accounts moved me towards the readers. My reflexive journal records that movement, and I report it so that you can weigh my reading with it in mind.

What This Means for the Writer

Accuracy will not protect you. Every participant believed the news, whichever version they read. If you judge a message only by whether it is correct, complete and clearly organised, you will miss what these readers were judging. Before you send anything that carries weight for a relationship, ask whether the people receiving it will feel you were talking to them.

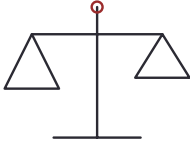
P16's test is a good one to borrow: is it written to the team, or for the team?

When the news is hard, the instinct is to add reassurance, thanks and encouragement until the message feels kind enough. For several participants, that extra layer was what made the generated message feel like management. Short, direct acknowledgement of the loss, a word about the actual work people did, and a clear next step read to them as respect. Long consolation read as a judgement that they could not cope. Trust your readers to take bad news as adults, and write in a way that shows you do.

Part of what made the human-written message feel addressed was that it named a colleague and promised to pass on the client's feedback. Whatever the study can and cannot separate, those are things only someone who knows the team could have written. If you use a tool to draft a message to people you work with, put these parts back by hand: the name, the shared history, the specific thing someone did, the promise you will actually keep.

You will usually not find out when this goes wrong. P6 never asked the supervisor the question. P8 did not tell the leader the message was insulting. P7 kept the view that the boss was not serious. Readers who feel written at rarely say so; they adjust their regard, their effort and sometimes their sense of who they answer to, and carry on. Do not treat an absence of complaints as evidence that your messages are landing.

Finally, keep P2 in mind. Some readers will read from your side of the desk and welcome the cushioned, careful message. You cannot write for every reader at once, but you can decide which relationship a given message is serving. Where its purpose is to inform, a tool can help and few readers will mind. Where its purpose is to stand with people through something hard, to give feedback on their work, or to acknowledge what they did, the readers in this study wanted the sender to turn and face them, and they could tell when that had been delegated.



CHAPTER FOUR

When Authenticity Matters, and When It Does Not

Whether AI authorship matters depends on the message's purpose, its channel and the writer's role.

Bad News Should Not Be Rushed

The message every group read was bad news, and P3 objected to the machine-generated version on grounds of timing:

we're not rushing to solution. We're not rushing to what's next. We are just being hit by this information that we have lost a bid that is probably life changing (P3)

A little later the participant said what the moment required: "at the end of the day, we are human beings. We need to 1st get down to how do we feel about this? And then now talk about what we will do next." The machine version moves quickly to what comes next. By its third paragraph it is telling the team: "While we cannot change the outcome, we can control how we respond." It adds that losing the client "presents an opportunity for us to reflect, learn, and continue delivering excellent work." Every sen-

tence is reasonable. Read on the afternoon the bid is lost, the sequence asks the team to be resilient before it has let them be disappointed.

Bad news, on P3's account, has a pace, and a message that arrives at the wrong speed fails whatever it contains. Only some messages have to meet that standard. A notice about a changed meeting room has no need to give its readers time to feel anything.

Seven of the sixteen participants spoke directly to the conditions under which a human hand matters, and those under which it does not. Their accounts leave the findings of the earlier chapters standing and mark out where they apply.

The Higher the Stakes, the More the Writer Matters

Set side by side, the participants' accounts apply a working rule that none of them stated in full: authorship matters in proportion to how much of the sender the occasion requires. A condolence needs its author. A routine notice does not. Most of the messages professionals write in a week sit between those two points, and the demand for a human hand rises and falls with three things participants kept returning to: what is at stake for the reader, the relationship between reader and sender, and the kind of message it is.

Stakes came first, and the message participants read sat near the top of the scale. Losing the largest client touches people's work, their standing and possibly their jobs, and P3's phrase "probably life changing" registers that weight. The occasions where participants found delegation least tolerable formed a recognisable group: bad news, condolence, feedback on someone's work and recognition of effort. In each, the reader is exposed in some way, and the sender has to show they have seen the reader's position.

Liu and colleagues, in an online experiment on AI-mediated email, found that trust in a writer fell when AI involvement was indicated, and that participants were least willing to accept AI assistance for emotionally sensitive messages such as condolences.¹ What the Nairobi discussions add

is the reasoning behind the reluctance, in the readers' own words, and a sense of how quickly it falls away as the occasion becomes ordinary.

Relationship was the second dimension. A colleague invited for a drink, P2 said, would confirm before the first sentence of the invitation was finished. P2's conclusion was general: "context of the message matters as to whether you will make it short or if it has to be a bit lengthy." The invitation and the bad news might pass between the same two people. What differs is how much of the relationship each message has to carry.

The kind of message was the third. Some messages exist to tell people something, some to be kept, some to get a thing done and some to hold a group in shape through a hard week. A message announcing a lost account does several of these at once. A message confirming that the invoices have been sent does only one, and participants did not expect it to do more.

The warranting principle helps explain why the demand moves. Walther and Parks proposed that people trust information about someone more when it is hard for that person to fake or control.² DeAndrea later stressed that readers actively weigh the cues in front of them, and that the same cue carries a different value depending on its source, its context and how easily it could have been produced.³ On a routine notice, the reader is not looking for evidence of the person, so its absence costs nothing. On a message of bad news, the reader is looking hard, and a text that offers only fluency, structure and courtesy, which any competent writer or system can produce, leaves them with nothing to hold.

Mayer and colleagues described three grounds on which we judge whether someone is trustworthy: ability, benevolence and integrity.⁴ Co-man and Cardon found that AI-assisted text holds up well on the first, being read as professional and competent, and loses ground on the other two, where readers see the sender as less sincere and less personally invested.⁵ The cost of a missing human hand therefore depends on how much a message leans on benevolence. A notice about a system upgrade leans almost entirely on ability: is the information right, and is it clear? A message

to a team that has just lost a client leans heavily on benevolence: does this person care what has happened to us?

Email Is for the Record

Several participants held, with good reason, that much of the writing that crosses a professional's desk owes the reader nothing personal at all. P12 put it this way:

I usually see emails is mainly for documentation purposes. If you want to have a real feel of how someone feels, calls are better (P12)

The participant gave the reason directly: "In case things go sideways, they have evidence that they can pull. So people will tend to be not to bring the authentic self when writing an e-mail." Asked how the machine-written message made the participant feel about its sender, P12 declined the premise: "I don't think too much about it because it's corporate at the end of the day." In the following session, P13 described the same function from the institutional side, noting that "one thing that is very key in our place of work is emails being used as accountability mechanisms."

On these accounts, email is an instrument of record before it is a place for relationship. It is where decisions are logged, where instructions are given in a form that can be produced later, and where people protect themselves against a future dispute. Expecting a personal voice in a document written to be filed would be a category mistake.

P9 reached a version of the same view by a more suspicious route, reading excess formality as a warning: "if an e-mail is filled too much. Uh, jargon and too much English. Then I feel [...] someone just trying to... Gets. Evidence, or to record something for purposes of future reference." A heavily formal email, read this way, is a message preparing for an argument. P9's own standard was practical and spare: "the most important thing is for me to act. Either you want me to do something, right, or you want me to know something." The participant also ranked media by how much feeling each can carry, and placed email near the bottom:

WhatsApps, chats, I always have raw emotions. e-mail sometimes is too formal (P9)

Sincerity, in P9's account, has somewhere else to live. If a colleague's real reaction to a setback reaches me in a WhatsApp message or a phone call, the formal email that follows can be as impersonal as it likes. It was never the place I would have looked.

The study's main findings are not a general case against AI-assisted writing. A minority of participants held, reasonably, that AI authorship barely matters if the message does its job. Coman and Cardon raise a related possibility: audiences in time-pressured professional settings may prefer the even, homogenised voice that AI assistance produces, because it takes less effort to pull the relevant information out of it.⁶ Both positions describe the other end of the same scale. In my own work as a communications specialist, a large share of what I write is meant to be accurate, consistent and forgettable, and its readers are well served when it is exactly that.

Most professional email confirms, schedules, forwards, reminds and records. Participants did not read such messages looking for their author, because they knew what they were reading and what it was for. For much of the working week, the demand for a human hand stays off.

Where You Work Shapes What You Expect

The sixteen participants were chosen to differ in where they worked: academic, corporate, institutional and community settings across Nairobi. Their expectations of a message followed their settings. Participants in procedural environments tested a message first against institutional form, and participants in close teams read informality as evidence of being close.

P2 drew a criterion from the structures the participant works within:

from the perspective of government, there is a structure. So at times you may be able to know whether the communication or the message is genuine when it follows a certain structure (P2)

Genuineness here means conformity to an expected shape. A departure has a practical consequence: “I would be hesitant to act based on that, but I’ll try to follow up and see that is this message really genuine.” The test produces a delay and an enquiry, with no particular feeling about the sender. P3 wanted to know whether the writer had been present. P2 wanted to know whether the message was legitimate, and in that setting legitimacy is carried by form.

This has a consequence for AI-assisted writing that is easy to miss. In a procedural setting, a machine-produced message that follows the expected structure may pass the reader’s first test more easily than a warm, irregular one written by hand. The idiosyncrasies that, elsewhere in this book, let readers recognise a colleague may count against a message where departures from form are a reason to check.

P1 read register as evidence about the workplace that produced it. Describing teams that “tend to work horizontally [...] like the boss and everyone else is sharing the same office space. There’s no walls”, the participant linked informality with physical and social closeness, and observed that people working this way “will try to identify more with Swahili rather than the formality.” In such a team, a formal email from the person at the next desk would sound odd, perhaps cold. P1 was clear that the formal version remains entirely appropriate in a ministry or in a mission abroad, and summed up: “the work environment has bigger influence on the kind of writing that we do.”

The two messages sit inside this pattern. The version I wrote is conversational, signed with a first name, and names a colleague whose pitch was strong. The machine version is formal, signed with initials, and thanks “each of you” for “hard work, commitment, and professionalism.” In a horizontal team of the kind P1 described, the second register signals distance. In a procedural environment of the kind P2 described, it might pass without comment.

Nkembuh, studying AI-mediated communication in remote professional collaboration, concluded that the relationship between AI medi-

ation and trust is moderated by transparency, familiarity and context.⁷ Wang and Adzharuddin, in a very different field, found trust in AI-generated content to be context-dependent and shaped by cultural familiarity.⁸ The accounts here show what that moderation consists of for these readers. The same absent author is a grievance in a message of bad news, an irrelevance in a channel kept for the record, and a reason to check before acting in a procedural environment.

Langmia and Haddad Nunez describe African communicative traditions as placing weight on relational accountability and on the embodied presence of the communicator.⁹ P1's remark about Swahili in close teams is one small sign of how that plays out in a Nairobi office: a shift of language and register marks a shift in relationship. It suggests that a tool which pulls every message towards one standard professional register will fit some workplaces better than others, and that writers in close, informal teams may notice the mismatch first.

IN PRACTICE: WHEN A DEVICE IS FINE

Every pattern that now reads as machine-made was used by people long before these tools existed. Speechwriters use the rule of three and good reporters open on a scene. Hunting every device out of a text leaves flat, choppy prose that reads as over-edited, which is a tell of its own. A device can stay if it passes three checks. It carries content: each item is true, the metaphor has a real second half, the contrast rules out a reading the reader would otherwise hold. It stands alone on the page, so it reads as a choice rather than a habit. And you would say it aloud to a colleague in those words.

Readers Who Do Not Care Who Wrote It

Some readers never reach the question at all. P11 handles messages in the order they arrive:

either via e-mail, WhatsApp, or a phone call. So whichever I see first or whichever comes first (P11)

For this participant, length is a barrier met before any content: “I get tired reading long emails usually [...] when you have too much on your plate and someone writes an essay, then it becomes difficult to spare time to read it.” When P11 does decide to give a message proper attention, the sender settles it: “depending on who it’s coming from, I would first pack it and finish up what I’m doing.” Whether a person or a machine wrote the message does not enter into it.

Walther’s early work on computer-mediated communication argued that people form impressions of each other through text as fully as they do face to face, given enough time and enough motivation.¹⁰ A reader triaging a full inbox has neither, so the slow reading that would surface a missing hand never takes place. The focus groups gave participants both: an hour, two texts, and a researcher asking them to look closely. That is a long way from P11’s ordinary day.

Under pressure, relationship works as a filter. A busy reader still decides which messages deserve to be set aside and read properly, and decides by who wrote them. Those few messages are where a writer’s choices about how to draft are most visible.

P12, who said more clearly than anyone that email is for documentation, was asked in the same session to judge the machine-generated message, and condemned it: the message “sounds lazy”, and its author, if a person wrote it, was “a lazy individual”.

I have not treated this as an inconsistency to be explained away. P12’s view of email is a position held about the channel, and it does not fully describe how P12 reads. A person can believe sincerely that email is a filing system and still feel the absence of a person in an email announcing a loss. Chapter 5 returns to what it means that so many participants who use these tools judged others for using them.

P12’s view of email, P9’s ranking of channels, P2’s test of form and P11’s triage all describe readers for whom, at least some of the time, AI authorship does not matter. In my reading they are applying the same rule as

P3, to occasions that ask less of the sender. Their divergence shows how far the rule stretches.

Everyone read a message of bad news, so the lower end of the scale rests on what participants told me about their daily work, not on anything I observed in the sessions.

What This Means for the Writer

The working rule participants applied can be put as a question to ask before drafting: how much of me does this occasion require? If the honest answer is very little, because the message records a decision, confirms an arrangement or passes on information the reader needs to act on, then AI help is unlikely to cost you anything with readers like these. If the answer is a great deal, because the message carries bad news, offers condolence, gives feedback on someone's work or recognises what people have done, the recommendation I drew from the study applies: where the purpose of a message is relational, write it yourself; where it is informational, generation can serve.

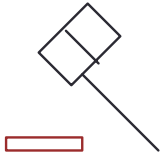
Three checks help with the harder cases in between. Consider the stakes: does this news touch the reader's work, standing or feelings? Consider the relationship: will this reader look for signs of you, or process the message in order of arrival? Consider the kind of message: is it written to be acted on and filed, or to hold people steady? A message that scores high on any of these deserves your own hand. One that scores low on all three can be drafted with help and checked for accuracy. When you are writing bad news, think also about pace. P3 objected to a message that moved to solutions before it had let people take in the loss, and that is something a writer controls whatever tool they use.

Know the setting your readers work in. In a procedural environment, a departure from the expected form can prompt a check rather than a warm response. In a close, informal team, the same formality can read as distance. AI tools tend to produce one register for every occasion, so be-

fore you send an assisted draft, ask whether it sounds like the place it is going to.

Use your channels deliberately. Several participants treated email as a record and expected to find the person elsewhere: in a call, in a chat, in a conversation. If that is how your workplace uses email, an impersonal email is acceptable only so long as people can still find you where they look for you. For organisations, if relational communication drifts into a channel nobody reads for a person, the consequences participants described elsewhere in this book are likely to follow. A team that sets norms about which occasions may be delegated, and keeps a place where people are expected to appear, is on firmer ground than one that tries to police the traces of the tools.

Finally, be careful with readers who tell you it does not matter. Some of them mean it, and for much of their working week they are right. P12, who held email to be a record, still judged an email about a loss as the work of a lazy sender. The safest assumption is that the demand for a human hand switches on with the occasion, whatever a reader says about the channel in general.



CHAPTER FIVE

Prosecuting From the Dock

People condemn AI writing in others while using it themselves.

ONE PARTICIPANT SPENT MUCH OF THEIR SESSION BUILDING A case against machine writing. P3 described colleagues whose filed documents had been flagged for AI, and drew a firm conclusion: “it feels like you’re being lazy when you’re using AI. So you have to come back. [...] And do it on your own.” Later, talking about a leader who has to deliver hard news, P3 asked that such a message come “from you, from your brain.”

Then, in their final turn of the evening, P3 said this:

So even I use a lot of AI. And I have just discovered it might not really connect or resonate to some other people (P3)

P3 did not propose to write more of their own messages. The remedy they offered was this:

I can add my personal touch, into it so that other people feel like I am also approaching it not from a mechanical point of view all the time (P3)

As a reader, P3 wanted to know that a message had come from the person who signed it. As a writer, P3 offered the impression that it had. Both positions were stated within seventeen turns of each other, and P3 gave no sign of noticing that they pulled in different directions. I have not tried to resolve that for them, because they did not resolve it themselves, and I am not sure anyone in the room, including me, could have done better.

Eleven of the sixteen participants contributed to this theme. People who judged others for using AI to write described, sometimes in the same breath, their own use of it. And when they did, the standard they applied to others changed shape.

I am not writing this from the gallery. I use these tools in my own work, and they helped me prepare this book. When I heard participants shifting the line from one place to another, I recognised the moves partly because I have made them myself. The title of this chapter describes all of us: we sit in the dock while we argue the prosecution's case. I mean it with sympathy, and I think the participants deserve it.

P3 arrived at the tension from the prosecution's side, condemning first and confessing later. P12 arrived from the other direction, offering heavy personal use of AI as the very thing that qualified them to judge. Their account of the tool: "once I learned to customize the prompts, to sound, to speak as if in my personality, it became quite easy for me." They gave the arrangement "a solid 80" for making their work simpler. Their verdict on the machine-written version of the lost-client message drew on that same experience:

Being an avid user of AI, from the second e-mail I would say it sounds lazy. [...] I find the author of e-mail 2 to be a lazy individual, if it was written by an individual (P12)

After I told the group how each text had been produced, P12 restated their position, saying that "with the correct prompt. Bringing AI on board works magic. from my perspective, so you can call me through AI, for lack of a better word." So P12 holds two sincere views at once: a mes-

sage from a thin instruction is lazy, and one from a good instruction is something P12 is happy to sign. For P12, the skill of the person giving the instructions decides which it is.

P3 sounded wary of AI and P12 enthusiastic. Each used it and condemned the machine-written message, and neither proposed writing the words themselves.

How People Excuse Their Own AI Use

Once participants had to square their judgement of others with their own practice, they kept the standard and moved it. The study found five places it went: to the prompt, to ownership of the content, to a final human pass, to how much of the text the tool produces, and to whether the tool's involvement can be seen at all.

The first place it went was the prompt. P16 had elsewhere compared a person's writing to their handwriting, a mark colleagues come to recognise without being taught. Later, P16 put the alternative view directly, and addressed it to me:

I can still use AI to produce a personally connecting message. And the secret, [MOD], as you know, is in the prompting (P16)

I read this as working against P16's own handwriting comparison, since a hand is the thing that cannot be commissioned. If a personal quality can survive being specified by one party and executed by another, the handwriting idea no longer holds. P16 did not seem troubled by this, and many readers will not be either.

P2 had already read the two messages in these terms before I disclosed anything. The human-written message was signed by a team leader called Siddondo, and P2 imagined a process in which the human intention runs all the way through: "The first one may have been done by Siddondo. Then I used a bit of AI to do the second one to bring a bit of the human feel." On this account nothing was lost, because the writer never left the message.

A second resolution replaces the question of who wrote the words with the question of who answers for them. P9 put this forward twice, once before the disclosure and once after, and their test was memory:

you sit down and you mean every word that. So you can actually remember what you wrote (P9)

The failure P9 wanted to detect was abdication, not automation:

I'll feel sort of offended if you even don't know what you sent me. AI sent it, you just send it. If I ask you a question about what you're saying, you don't remember, I'll feel offended (P9)

Where the sender keeps that relation to the words, P9 had no objection: "it told you something, you own it. Is fine." This is the most practical idea in the theme, because it gives a writer something to check before sending. The objection had been about where the words came from, and P9's test asks whether the sender can stand behind them.

A third resolution sits close to the second. P11 described a clear division of labour, in which the thinking belongs to P11 and the drafting does not: "you have a couple of thoughts together, but you don't have the time to think through and formulate something [...] So you add up my points in place and then send it to the tool." The final step restores P11's authorship, as they see it: "Then, I'll come back to just do the final refinement to ensure that it's using my tone." By the close of the session P11 had set the authorship question aside altogether: "whether one has used AI or they have written by from the heart [...] it's as long as the message has been passed and has been passed correctly, I think that's the baseline." I return to P11 later.

P13 turned the standard into a matter of competence. AI is "just any other technological advancement, just as basic as the innovation of phones [...] So this is just another stage." What the user owes is verification: "if you don't take time and really look deep into what has been present to

you by AI, you might be communicating a lot of misinformation.” Without the right inputs, it is “garbage in, garbage out.”

Two participants took a position of degree. P15 located the fault in quantity: “the problem becomes when we over utilize AI, we let it now speak with its own language and it loses the human part of it.” Used properly, P15 allowed, “it can be a very good tool.” P14 compared the tool to ordinary delegation, “like giving your secretary a job to do a certain, to write that communication,” and added that a skilled user leaves no trace: “people who can use that AI, and you most probably will not realize. So it also looks at their skill in using that particular tool.”

That remark leads to the last place the line moved. If a competent user’s AI involvement cannot be seen, then whether it can be seen starts to look like the thing that matters.

The Zero Per Cent Rule

P16 described an institution that works on that assumption. Speeches prepared for a national principal are checked before release, and the target is absolute:

normally we want to aim for a 0% AI similarity to ensure that really there’s no room for any possible sighting of [...] the president read an AI generated speech (P16)

The rule aims at a speech in which the tool cannot be detected, which is a different thing from a speech in which the speaker is present; a speech could pass the check and still sound like nobody in particular. And the rule did not come from readers. P16 said that “the move was not really informed by any feedback we got on our communication.” No audience had complained. The rule arose to manage a risk of embarrassment: a speech that had been, in P16’s vivid phrase, “fully AI marinated.”

I do not offer this as a criticism: a team writing for a senior public figure has good reason to avoid a headline about machine-written speeches.

What interests me is how naturally the standard moved from authenticity to detectability, without anyone seeming to decide that it should.

The published research suggests the target is harder to meet than it looks. Masrour and colleagues tested nineteen tools sold to make AI text read as human, and found that they swapped one set of machine patterns for another and did not restore the features of human writing.¹ An industry has grown up around the same goal P16's organisation set itself, and it manages the trace while leaving the reader's experience much where it was.

Others in the study described the same logic at the level of the individual. One participant described users who have learned to hide the marks that would give the tool away. After I revealed how the two messages had been made, P8 moved from objecting to any use of AI to objecting only when the use is noticeable. Several participants, on learning which message the machine had written, revised the terms of their objection rather than its strength. One who had called it inauthentic and lazy afterwards allowed that polishing a reader could not see would give no difficulty. For some, finding out turned a prohibition into a threshold: the tool was acceptable up to the point at which it showed.

I do not think this makes anyone a hypocrite. These are people trying to find a rule they can live by from both sides of the inbox. The difficulty is that the rule they find most easily, keeping the tool out of sight, is the one that does least for the reader.

Should You Say You Used AI?

If the line keeps moving towards invisibility, the obvious answer is disclosure. The research on this is less comforting than it sounds.

Schilke and Reimann, across thirteen experiments, found that people who disclose their use of AI are trusted less than those who do not, because they seem less legitimate, whether the disclosure is voluntary or required.² It was worse when someone else exposed the AI use than when the sender admitted it. Coman and Cardon, surveying close to nine hun-

dred working professionals, found that AI-written workplace messages were rated as professional and effective, but that ratings of sincerity and caring dropped once AI authorship was known.³ Zhu and Molnar found that senders known to have used AI were rated as lazier and less sincere, while undisclosed AI messages were rated as well as human ones and readers rarely suspected anything.⁴

Sahebi and colleagues call this a disclosure paradox: readers believe AI use ought to be disclosed, and they penalise the message when it is.⁵ If being found out is worse than admitting it, and admitting it is worse than saying nothing, then saying nothing and hiding the trace is the rational choice for a writer, and P16's zero per cent rule shows the same choice made by an institution. It is a poor outcome for everyone who reads.

Drozdz and Söilen, studying online reviews, found that text labelled as AI-generated was trusted least, while text labelled as AI-assisted was judged more favourably, and that the difference ran through the reader's sense of authenticity.⁶ What mattered was whether the label told readers a person was still involved. That fits closely with what my participants were reaching for when they moved the line to ownership and to the final pass.

I should be open about my own stake here. After the disclosure in the first session, the group's responses did not match what I expected, and I wrote in my journal that night that I felt bad. I had brought an expectation, built from my own practice, of how disclosure would land, and the room declined to confirm it. I recorded that feeling as information about me and not about them, and I have tried to keep it out of how I report what they said.

Only one participant accepted the difficulty this chapter describes as their own. Shortly before P14 spoke, I had used the phrase "double standard" as moderator, so what follows is P14 reflecting on a term I offered, not an independent naming of it:

honestly, it is a double standard [...] we don't own the knowledge and understanding of how to use AI. All of us are learning every day (P14)

P14 did not withdraw their reading of the machine-written message. They withdrew their entitlement to judge the person who produced it, on the grounds that the judge is still learning too. I find this the most generous response in the study.

I do not want to treat the double standard as a gotcha. People often hold others to standards they struggle to meet, and these participants were not hiding their own use; they disclosed it themselves, in front of colleagues they had just met.

IN PRACTICE: DRAFTING WITH A TOOL

If you draft with an AI tool, prevention costs less than repair. At the start, give the tool a list of the phrasings you do not want, ask for plain declarative prose, and ask it to flag any sentence it cannot support with a fact from your own material. Then edit the result yourself, because over a long draft the tool drifts back to its defaults. Write for the reader, not for a detector. Detection tools are unreliable, and a text written to slip past them is rarely a better text for the person who has to read it.

Two Exceptions

Two participants complicate the picture.

P10's concern was for a third person's voice. P10 described preparing a speech for a principal and finding that the tool produced something "too far from what how authentic self is." The draft could be adjusted, but "sometimes it still comes back and the speaker is like, this is not me." P10 did not claim to have solved this: "So I think for me, I've struggled with that." Where other participants were negotiating a standard for themselves, P10 was answerable for somebody else's authenticity. The speaker can say "this is not me" with authority, and the writer has to find out what "me" sounds like for someone else while the tool keeps pulling towards a voice that belongs to no one.

P11's account runs the other way. P11 was the only participant who described being harmed by another person's use of AI. A project manager

put P11's own response through an AI tool, and the reply that came back read to P11 as a personal attack, with senior colleagues copied in. That experience sits directly against P11's later position that the only baseline is whether the message has been passed correctly, and P11 gave no sign of noticing the tension.

I have left this case open. P11 may have been agreeing with the direction the group had taken by the end of the session, which happens in focus groups. Or P11 may have moved the standard to the human pass before sending, distinguishing their own careful use from the project manager's careless one. The transcript does not settle which. It does show is that someone who has received a bad AI message can still, a short time later, describe a rule that would have allowed it.

What the Double Standard Shows

None of the participants who condemned AI writing in others resolved the tension by proposing to write more themselves. They resolved it by moving the criterion: to the prompt, to ownership, to the final pass, to proper dosage, or to whether the trace can be seen. Effort remained the measure by which they judged others. The effort they put in to meet that measure was increasingly directed at the tool rather than at the text.

The study's conclusion is that the test of authenticity is moving away from authorship and towards accountability and undetectability. Accountability, in P9's sense, keeps a person attached to a message. They meant it, they can remember it, they can answer a question about it. Undetectability keeps a person attached only to the appearance of a message. The second invites concealment, which is what the current incentives reward.

This shift is happening without deliberation. Individual writers moved their own line one reasonable step at a time, while continuing to judge others by the older standard, in which a message should come "from you, from your brain." That gap between how we judge and how we work

is the double standard, and I think most professionals who use these tools are living inside it.

Walther, whose warranting principle runs through this book, argued that people trust cues that are hard for a sender to fake.⁷ The participants were, in effect, searching for a version of that principle that could survive their own use of the tools. Ownership is one candidate: a sender who can explain and defend a message has given the reader something difficult to fake. Detectability is a poor candidate, because it measures only whether the faking was done well.

What This Means for the Writer

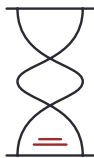
The first thing I would say to any writer who uses these tools, myself included, is to admit the double standard. Most of us judge a colleague's machine-written message more harshly than our own, because we know the intention behind ours and cannot see theirs. That also means the reader of your message is likely to apply the older, stricter standard to you, even if they use the same tools themselves.

Before you send a message that a tool helped to draft, apply P9's test: ask whether you mean every word, whether you will remember what it said, and whether you could answer a question about it without going back to check. If the answer is no, the message is not yet yours, whatever its origin. This test does not forbid the tool. It asks that you remain the person accountable for the message, which is what a reader is entitled to expect.

Be wary of making invisibility your goal. The research on humanising tools, and P16's account of the zero per cent rule, both suggest that this manages the trace without restoring the person. The better use of editing time is to put back what only you could supply: the name of the colleague who did the work, the fact you know about this team, the plain sentence you would have said aloud. Those details are hard to fake, and they are what readers in this study were looking for.

On disclosure, there is no free choice. Telling a reader that you used AI may cost you some of their trust. Hiding it carries its own risk if someone else finds out. A sensible working rule is to disclose where the relationship makes concealment the greater danger, which usually means with people you work closely with and will keep working with. Where you do disclose, describe what you did in a way that shows you were present: that you drafted with help, checked the facts and wrote the parts that matter yourself.

Finally, if you write for other people, as P10 does, take their objection seriously when they say a draft is not them. The tool will keep pulling towards a voice that belongs to nobody, and the person whose name is on the message is the only one who can say when it has gone too far. That is slower than accepting the first draft, but it is the only way the speaker's audience will hear the speaker.



CHAPTER SIX

A Warrant With an Expiry Date

Telling a person from a machine takes familiarity or a comparison, and the clues may not last.

IN ONE OF THE SESSIONS, AFTER THE GROUP HAD SPENT SOME time comparing the two messages, P13 said something that cut under the whole exercise. The others had been listing what gave the machine away: the length, the structure, the polish, the missing name. P13 agreed that the differences were there and then asked where they came from.

I think the only reason why we are able to tell the difference is because we have the two. If I receive the first one, it communicates the same thing as the second message

I had designed the session around two texts on one screen, and here was a participant pointing out, politely and from inside that design, that the design was doing some of the work. P13 then applied the point to the people they actually write for: “majority of the people we communicate to

might not be able to pick that at a glance. So it's easier for us because now we have what to compare with."

P13's remark asks what the clues of the first chapter are worth to a reader who receives one message, from someone they do not know, with nothing beside it. Only five participants, from three of the four sessions, took up that question. I give their accounts a full chapter because they concern how far readers can trust their own eyes, and because what they said reshaped the theory the whole study rests on.

Knowing the Sender

Look again at what each clue from the first chapter needs. P10's colleagues recognise their messages by the dots they use a lot, which works only for people who know the dots. P12 described a colleague who never greets, so that a sudden burst of courtesy gives the machine away; you can only see a change if you know what came before. P16's classmates knew each other's handwriting from years of seeing it. P13's story of the attempted fraud collapsed on a single fact, that the colleague asking for money had no spouse.

Every one of these clues is a comparison. The reader holds a picture of the sender, built over months or years of ordinary contact, and measures the new message against it. A grammatical slip or a sudden politeness means nothing on its own. Each means something against a background, and the background belongs to the reader.

P6 saw what happens when that background is missing. Late in the session, after I had told the group how each text was produced, the participant described a writer who has become skilled with the tools:

you've learned the art of using AI properly to refine a message, look really refined, hiding the marks of the AI itself. It's like faking it for so long because they have never known you before

P6 built a scene around it, a new manager arriving in an unfamiliar team where everyone writes beautifully: "if you are a new manager in a new

space and you found people are writing very nice articles [...] would you argue that I have a very fine team?” The manager is paying attention, and concludes that the team is strong, because they have no earlier version of these people to compare the polish against. To them it looks like competence.

Many professional readers are in that position. A recruiter reading a cover letter, a programme officer reading a proposal from an organisation they have not funded before, a client reading the first email from a new agency: none of them has a baseline for the writer. What they see is fluency, structure and courtesy, which any competent writer or any current system can produce, and they read those qualities as information about the person.

P6 made this point after learning which message the machine had written, when they had nothing further to gain from it, and it sat awkwardly with their earlier defence of re-voicing a message with the tool. They did not reconcile the two, and I have left them as they stand.

P13 pushed the same point further, sorting the people they write for by how closely they read:

to what extent are the masses familiar with this? Myself, I might be familiar, but the people I work with, the staff that respond to me, my peers and also my managers, to what extent will they pay attention to those cues

For P13 this settled their own position. Whatever the two messages differed in, their response to the news would be the same: “So the emotions might Be just the same to my context.” In the first chapter I described P13’s warrant as institutional: the house font, the copy list, the patterns that grow up among colleagues. A stranger’s email gives a reader like that nothing to check.

P16 accepted the point in passing, remarking that “[P13] is simply telling us the difference is the same. Which is fantastic,” before moving on to their own question about how a message makes its reader feel. P15 had the last substantive word:

for those of us who are aware of AI, we know we can tell when we see an e-mail and for the rest who are not aware, probably, I don't know how they probably think about the whole thing

P15 was confident about the group's own ability and unsure about everyone else's, and said so without embarrassment.

Asking ChatGPT Why It Writes That Way

In one session, P8 asked for the floor: "Dekker, Dekker. Sorry, sorry, sorry for the interruption. Do I have permission to speak?" While the rest of us had been discussing the two messages, the participant had been running an experiment of their own.

Just an interesting thing, yeah? So, the first version of that e-mail, I fed it, I fed it, I fed part of it to ChatGPT, like as we were speaking, and I asked it to, you know, to polish it, to improve, review and improve it

The first version was my own short email, with a colleague named in it. What came back was much longer, and P8 read the opening aloud:

And it has given me around 6 paragraphs, yeah. And it goes, I'm sorry to share that Sankara has informed us that they will be appointing another agency following the recent review of the account. I know this is disappointing news, particularly for everyone who has invested significant time, energy and care in the account and the application process

The passage is fluent, considerate and correct. It also shows, within two sentences, several of the features the group had spent the previous half hour naming as signs of the machine: the softened announcement, the effort listed in three parts, "time, energy and care", and a phrase that covers everyone so that no one in particular has to be named.

What interested P8 most was the tool's short account of what it had done and why:

the interesting thing I got is how it tries to justify that new draft, yeah? And it says that this version avoids sounding overly casual. Acknowledges the team's

disappointment and gives a clear next step without creating unnecessary alarm

The implication reached the participant halfway through a sentence:

Yeah, and yet it, and yet, and yet it is the kind of casual touch in the e-mail that me and I think one or two other people were pointing to as the aspect of authenticity and, you know, the personal touch, yeah

This is a single, uncontrolled episode, one participant using one tool on part of one text, and I offer it as an illustration, not as evidence of how these systems behave in general. Three things in it stand out.

First, the tool stated its own aim: to avoid sounding overly casual, which the draft did. Second, the quality the tool treated as a fault was the quality this group had already, on their own, named as the evidence of a person. Casualness was what P6 had chosen over correctness when picking the flawed message. It was what P4 found survived being read aloud. It was in the writer P5 described, who mentions a thing and moves on without explaining it.

Third, the removal was presented as an improvement, in the tone of a helpful colleague. A user who accepted the new draft would have been told exactly what had been taken out without being in any position to see it as a loss. In this instance, the tool's idea of a better email ran directly against the clues readers use to find the writer, and it did so in the open.

P8 later said of the second message that "the scaffolding I think that you are talking about is a bit too much". The observation was P8's own, but the word scaffolding was mine, offered earlier in response to P5.

When Polish Becomes Normal

Several participants expected the clues to change sides. P5 put it in terms of what readers will come to expect:

it is becoming normal to use AI to the extent that now if you are writing without using AI, people will now feel that you are the one who is now, your English is not well polished

Today, for P5 and others, a rough edge in a message is a sign that a person wrote it. If polished prose becomes the default every reader expects, the same rough edge starts pointing to a weak writer. The evidence stays on the page while its meaning reverses.

P8 raised the same expectation about the things participants had described as most local, the idiom, register and voice that place a writer in a particular culture: “What if we get to a point where AI figures out these all these all unique voice thing, these all cultural nuance and context thing? [...] Because these are machines and they are also learning very fast.” P1 held a version of this view from the start, refusing to convict either message on any single clue because “AI can do both documents”, and said so before I had disclosed anything about how either text was produced.

Outside evidence suggests the tools already do some of what these participants feared, though not well. A study of nineteen tools sold to make machine text read as human found that they often got past detection software but did not restore the features of human writing. They swapped one set of machine patterns for another.¹ Other work points away from P8’s worry. Writing assistance has been found to move text towards Western conventions of style and to wear away cultural markers,² and machine rewriting has been found to keep most of a text’s meaning while removing the signals a reader uses to infer who wrote it.³ For now, the tools are more likely to erase a local voice than to imitate one. Either way the reader is left with fewer clues that can be trusted.

There is a third route by which the clues expire, the one P6 described: writers learn which features give them away and take them out. Researchers studying disclosure have described a paradox in which readers say AI use should be declared and then penalise messages when it is, which gives

writers a reason not to declare it.⁴ The writers P6 described, who have learned to hide the marks of the tool, are that incentive in practice.

For one participant, the disclosure at the end of the session changed everything they had read before. P13 described the effect as it happened:

the script that you've given has really got me thinking. And I was here just trying to figure out how much AI generated emails have I consumed

The participant said they would now “pay a little bit more attention and pick out the emotion that comes from the messages”. P13’s case is divided: their argument that telling the difference depends on a comparison survived learning which text was which, but their habits as a reader did not, since they left going back over messages they had accepted without question.

P8 moved the other way. Having called AI-assisted writing inauthentic and lazy earlier in the session, P8 allowed after the disclosure that where the tool had “polished the language here and there without adding so much verbosity in a way that I barely or not notice at all, then I would have no problem”. The objection had become a threshold set by visibility, and P8 made the move without remarking on it. Two readers met the same news in opposite directions: one became more watchful, the other decided that what cannot be seen does not matter.

IN PRACTICE: THE DASH, AND OTHER DATED CLUES

The long dash is the clue general readers cite most often, and the ChatGPT message in this study used one. Replace each with a comma, colon, semicolon or brackets, depending on the sentence: “A significant majority—71%—reported” becomes “A significant majority (71%) reported”. Set apostrophes and quotation marks as curly pairs throughout, since mixed straight and curly marks often show where text was pasted from a chat window. Treat lists of clues like this one as a snapshot. As this chapter shows, tools learn the clues and readers move on to new ones. The material that keeps its value is the detail only you could have supplied.

Trust Depends on Knowing the Writer

Walther found that people forming impressions of each other through text adapt to the missing face and voice by reading more into whatever cues the text still carries.⁵ With Parks he then asked how a reader decides which of those cues to believe, given that a writer online can present themselves however they like. Their answer was the warranting principle: we give more weight to information the person could not easily have made up, and less to information they control.⁶ A colleague's account of you counts for more than your own description of yourself, because you did not write it.

DeAndrea later asked what makes a piece of information hard to fake, and identified two things. One is effort, meaning what it would cost to produce the information. The other is access, meaning whether producing it requires something only certain people have.⁷ A detail of a colleague's family life is cheap for someone close to them and almost impossible for anyone else, which is why the fraud P13 described failed.

My study proposed one addition. A writer's small involuntary habits, the dots, the dropped pronoun, the missing greeting, the slip left uncorrected, are hard to fake in both of DeAndrea's senses. They come from a particular person's history with language, they cost that person nothing, and nobody else can produce exactly the right ones. If that is so, those habits work as a warrant: they tell the reader that a particular person wrote this. A message stripped of them, fluent and correct and unremarkable, offers no such evidence. I called that condition warranting failure, and the first chapter describes what it looks like from the reader's side.

The participants showed me that I had put the warrant in the wrong place. The warranting literature treats warranting value as a property of the information: a cue is hard to fake or it is not. My own proposal had followed it, treating idiosyncrasy as something a text carries. The accounts of P6 and P13 do not fit that picture. P6's new manager reads a polished message from a writer who has hidden every trace of the tool, and finds nothing wrong, because they have no history with the writer. P13 points

out that the group could tell the two texts apart because they had both. In each case, the same message carries a warrant for one reader and none for another.

So the warrant lives in the relation between a text and a reader equipped to read it. A writer's habits are evidence only to someone who knows the habits. The dots P10 uses are a signature to colleagues who have read P10 for years and a meaningless quirk to a stranger. This is consistent with DeAndrea's view of warranting as a process of perception, in which readers weigh cues for themselves.⁸ It goes one step further, because it makes the reader's prior acquaintance a condition of the cue existing at all. Without that acquaintance, there is no clue to notice in the first place.

I think this is the most important thing the study adds to the theory it borrowed, and it came from the participants. I had proposed that idiosyncrasy works as a warrant. They showed me where that proposal stops.

I had anticipated part of this problem, drawing on research which holds that people with no other information about each other pay closer attention to whatever cues remain.⁹ P6's new manager does pay attention, and reads refinement as evidence of a capable team. What is missing is the expectation that would make the polish look strange. With a stranger, then, a writer's rough edges show that a person is present only to a reader who still expects people to write with rough edges. As polished machine text becomes the normal register of professional life, that expectation wears away, as P5 forecast.

Here I part company with some of the participants, who expected the ability to tell person from machine to disappear. On my reading of their own accounts, the link between a message and its sender does not vanish. It moves towards familiarity with a known sender, where P10's dots and P12's colleague's missing greeting still work, and towards the shared conventions of an institution, which is where P13 had placed their trust all along. A reader who knows the writer, or knows how messages in their organisation are supposed to look, still has something to check against. A reader with neither has very little.

The research on detection supports the participants' caution. In six experiments with 4,600 people, Jakesch and colleagues found that readers could not reliably tell machine-written self-descriptions from human ones. They relied on rules of thumb, such as first-person pronouns, informal contractions and mentions of family, that machine text displayed just as often.¹⁰ A study of German student theses found that human readers identified machine text 57 per cent of the time, a little better than chance, and below 20 per cent when the machine text was written in a professional register.¹¹

Participants in this study reached for clues of the kind Jakesch found unreliable, a dropped pronoun, a contraction, a proper name, and with both texts in front of them they were generally right about which was which. As Chapter 1 noted, the pairing may have made the difference easy to see, and the clues may work less well in an ordinary inbox.

Cardon and Coman also found that the trust cost of AI-assisted writing does not depend on the reader consciously spotting the tool.¹² If the clues expire, readers lose the means of explaining their unease, and the cost to the writer may remain all the same.

What Changes When You See Only One Message

P13's objection is also a fair criticism of my design. Participants read both messages in the same sitting, a comparison that a person reading one message in their inbox does not have, and the differences they named, in length, structure, warmth and a colleague's name, are what a side-by-side reading makes easy to see. The study does not show that these readers would identify a single machine-written message arriving on its own, and P13, P6 and P15 each suggested that they would not always manage it.

The two texts also differed in more than authorship. My version named a colleague, promised to share the client's feedback and wished the team a restful weekend. The machine's version had none of those things and added extended thanks and encouragement of its own. Some of what

participants responded to may be those differences of content, which the pair cannot separate from authorship.

Sixteen professionals in one city, most of whom used these tools themselves, are not a cross-section of anyone's workforce, and their accounts are not measures of how common any reaction is. Nothing in this study claims that readers can generally detect machine text. I take the point P6 and P13 made as a boundary on everything else in the book: the clues of the first chapter are real, readers use them, and they work best for readers who already know the writer or have something to compare against. The participants who made it were describing the conditions of their own competence, more frankly than the design required.

What This Means for the Writer

If you write to people who know you, your habits still carry weight. Your colleagues have a picture of how you write, and they will notice when a message departs from it, including when it departs upward into polish you do not usually show. Where you use a tool on a message to people who know you, read the result against your usual voice before you send it, and put back the turns of phrase that are yours.

If you write to people who do not know you, your habits cannot vouch for you, because the reader has nothing to measure them against. P6's new manager will read your polish as competence and learn nothing about you. What still carries weight with a stranger is material only you could have supplied: the name of the person who did the work, the specific detail of the situation, the commitment that ties you to a next step. These are cheap for you and hard for anyone else to invent.

Be careful with what the tools call improvement. A tool asked to polish a message may take out the features your readers use to find you, and tell you so in terms that sound like good advice. Look at what it took out before you accept what it put in. The casual line, the named colleague and the short sentence that states a point without explaining it may be the parts doing the most work.

Expect the ground to keep moving. Participants expected polished prose to become the norm, and some expected rough English to be read as weakness. If that happens, small errors will stop helping you with readers who do not know you. Your relationships and your organisation's shared ways of writing will matter more, because those are where a reader can still check a message against something. Letting colleagues come to know how you write, over time, gives them the baseline this chapter shows they need.



CHAPTER SEVEN

What Writers, Teams and Tool Makers Can Do

What the Six Findings Add Up To

Sixteen professionals in Nairobi read the same piece of bad news twice. I wrote one version, and ChatGPT produced the other from a prompt carrying the same facts. Everyone believed the news. What divided the two messages in their eyes was something the facts could not carry.

Readers looked for a person in the text. They found one in small, unplanned habits and in details only that sender could have known, and they found the machine in excess: more length, more structure, more reassurance than the occasion needed. Participants agreed that the machine-written message was correct, complete and courteous, and still found it wanting, because what they wanted was evidence of who had written it and how closely that person had attended to them. Where that evidence was missing, readers went on believing the message but no longer felt addressed by it. Their regard for the sender fell, and one participant described feeling less part of the team the message was written to.

The finding I take most seriously is that some readers who concluded that no effort had been spent described spending less themselves: reading only for the action items, replying with a stock phrase, or in one case putting less into the next piece of work. This rests on four people's accounts, and I offer it as a proposition to be tested, not a measured effect. If it holds, the cost of a delegated message falls on the work as well as on the relationship, and it goes unseen, because a reader who withdraws attention has no reason to say so.

The other findings set its limits. Bad news, feedback and recognition call for a person, while a routine notice in a channel kept for the record owes the reader nothing personal. The people who judged machine writing most harshly often used the tools themselves, and their standard shifted as they talked, away from who typed the words and towards who stands behind them, and in some cases towards whether the tool could be seen at all. The clues readers relied on worked only for readers who knew the sender or had a second text to compare, and participants expected those clues to stop working as tools learn them.

At work, then, a message carries information and also a record of how much the sender attended to the person receiving it. Tools can supply the first at almost no cost and imitate the surface of the second, but readers who notice the imitation respond to it. What follows turns that argument into guidance for people who write, the organisations they write in, and the people who design and govern writing tools. It comes from sixteen readers in one city, reading one kind of message with a comparison to hand, and I offer it as a set of considerations for settings like theirs.

FIGURE 4 Six findings on one page

The Hand in the Text Readers recognise a person by small habits and details only that person could know, and the machine by excess.	The Ledger of Effort Effort is judged by attention to the reader and the situation, not by length or time spent.
Written to Me, or Written at Me Readers can accept what a message says while withholding trust from the person who sent it.	When Authenticity Matters The demand for a human hand rises with the stakes and the relationship, and falls for routine messages.
Prosecuting From the Dock People judge AI use in others while using it themselves, and keep moving the line.	A Warrant With an Expiry Date The clues work only for a reader who knows the sender, and may fade as tools learn them.

For People Who Write

Most readers of this book, like most of my participants, already use these tools, and nothing in the study suggests that using them is the problem in itself. What it does suggest is where a writer’s own hand still counts.

Keep the particulars. The features that convinced participants were ones a machine could not supply from a brief instruction: a colleague’s name attached to what that colleague did, a reference to a shared history, an acknowledgement of what is not yet known. If you draft with assistance, these are the elements to restore by hand. A tool given the facts will write about the situation; only you know which person carried the work, what the team has been through, and what you cannot yet promise. Warranting theory explains why such details carry weight: readers trust information that would be costly or impossible for anyone else to produce.¹

Match the length and form of a message to its occasion. Participants read excess as a failure of care. A long message with headings, reassurance and a closing summary struck them as written for no one in particular.

Relevance theory describes communication as a bargain in which the reader's effort should be repaid by what the message gives them, and a padded message breaks that bargain from both ends: it cost the sender little and costs the reader more.² In a message of bad news, one participant wanted the sender to slow down and let people absorb what had happened before turning to next steps. Another read the brevity of the human-written message as a sign that the sender was affected. When a tool returns a draft twice the length you would have written, the extra words are usually the part to cut.

Leave your habits in. Participants recognised colleagues by small things: a punctuation mark they always used, a phrase they repeated, even their errors. Some read any departure from a known pattern, even an improvement, as a sign that someone else had written the text. This is no reason to add mistakes; participants themselves pointed out that a system can be told to insert errors, and a manufactured flaw is one more imitation. It is a reason to be cautious about accepting every correction a tool offers on a message to people who know you. The rhythm of your sentences, the word you habitually reach for, the Kiswahili word that says what your professional English cannot, may be what tells your colleagues it is you. Four participants located the mark of a person in local idiom, register and voice, and one described a colleague's shift into a borrowed register as a betrayal of that colleague's own voice.

Own what you send. Several participants proposed a standard that does not depend on who typed the words. P9, quoted in Chapter 5, put it as a test of memory: a sender who cannot answer a question about their own message has given offence. Read the draft as if the recipient will quote any sentence of it back to you next week and ask what you meant. If there is a sentence you could not explain or would not defend, the message is not yet yours, whoever drafted it.

Reserve the occasions that carry relationship. Bad news, condolence, feedback on someone's work and recognition of effort were the cases where participants found delegation least tolerable. One participant, after

discovering that a supervisor's long written feedback had been copied from ChatGPT, privately concluded that the supervisor had lost authority over them, and never raised it. Where the purpose of a message is relational, write it yourself; where it is informational, generation can serve. Most messages are some of each, and then the rule applies to their parts. A tool can lay out the dates, figures and actions in a project update. The sentence thanking the person who stayed late to finish it should come from you.

Apply to your own writing the test you apply to others. Chapter 5 described participants who condemned machine writing in colleagues and then described their own use of it, sometimes within minutes. I count myself among them, since I use these tools in my own practice. If you would take a colleague's long, smooth note of sympathy as a sign that they could not spare the time to write it, your reader will probably take yours the same way.

Consider disclosure carefully. When participants learned which text the machine had produced, most changed the terms of their objection more than its strength, and one moved from objecting to any AI use to objecting only when the use was noticeable. The wider research reports what its authors call a disclosure paradox: readers believe AI use should be disclosed, and then penalise the message when it is.³ Other experiments have found that disclosing AI use costs a communicator trust, and that the cost is greater still when someone else exposes the use.⁴ Concealment is a bet that the reader will never find out, and in a lasting working relationship, readers often do. My recommendation is to disclose where a working relationship makes concealment the greater risk, and to accept that the choice carries a cost either way.

For anyone who wants something to keep beside the keyboard, these habits reduce to four questions, each taken from what participants looked for:

- Does this message contain something only I could have written: a name, a shared fact, an admission of what I do not yet know?

- Is it longer, more structured or more reassuring than this occasion needs?
- Is its purpose relational or informational, and did I write the relational parts myself?
- Could I answer a question about every sentence in it?

IN PRACTICE: SAY IT ACROSS A TABLE

One test catches most of the habits described in these notes. Could you say the sentence aloud to a colleague sitting across a table from you, in those words, without sounding as if you were reading? If you could not, it is probably not your sentence. A second test helps with dense prose: if someone asked you to explain a sentence, would you use the same words? If you would have to translate yourself, write down the translation.

For Teams and Organisations

Organisations shape how their people use these tools through what they measure, permit and reward, often without deciding anything deliberately. The recommendations here are about making those decisions on purpose.

Count attention as well as time. Organisations tend to assess writing tools by the time they save at the sending end. Participants described a cost at the receiving end. Some read less closely, or replied more thinly, when they judged a message to have been generated. P7, quoted in Chapter 2, said they might put less effort into the next pitch because the boss had put none into giving feedback.

These are reported actions and intentions from four participants, not observed effects on output. They are consistent, though, with experimental work finding that perceived AI use reduces readers' willingness to take up and use what a message tells them.⁵ An organisation weighing the time a manager saves by generating a team update should also ask what happens to the attention the team gives it, and can put the question to staff

directly: how do you read messages you suspect were generated, and what do you do differently? The answers will say more about the tool's real cost than a count of hours saved.

Do not treat detectability as authenticity. P16 described an institution where speeches were checked before release, with the stated aim that “normally we want to aim for a 0% AI similarity”. The rule had not been prompted by any complaint from the audiences. A standard of that kind manages the visibility of the tool and leaves the relationship between speaker and audience where it was. It also tends to defeat itself: tools designed to make machine text pass as human have been found to replace one set of machine patterns with another.⁶ Most important, it changes what staff are asked to do, from writing as the speaker to hiding the tool. This was a single account, but it names a substitution that can happen without anyone choosing it.

Set norms rather than police traces. The productive question is what may be delegated and on what occasions, which a team can discuss in a meeting. Asking what can be detected invites concealment and an arms race with the tools. A norm can be simple. It might say that announcements of bad news, performance feedback, condolences and thanks are written by the person whose name is on them, and that reports, minutes, routine notices and procedural correspondence may be drafted with assistance, provided the sender has read, checked and can answer for the result. Such a norm also gives permission. Several participants described email as a channel kept for the record, where the right copy list and expected structure matter more than a personal voice. A norm that says so frees people from performing presence in messages that do not need it.

Protect the channels where people are expected to appear. Several participants held that relational communication lives somewhere other than email: in calls, in messaging, in person. In practice this argues for delivering serious news in a call or meeting where the team can hear the person who carries it, with the written message as the record that follows.

Expect the costs to concentrate on people with authority, and to go unreported. The participant whose supervisor's feedback turned out to be copied did not complain. They drew a private conclusion and kept working. This worries me most for managers, because it produces no signal. Research on trust in organisations distinguishes a person's ability from their benevolence and integrity, and studies of AI-assisted workplace writing find that the damage falls on the second pair while judgements of competence largely hold.^{7 8} A manager who is told their messages are clear and professional may be hearing about the one quality that was never at risk.

Make disclosure safe before requiring it. If an organisation expects staff to declare AI use and then judges declared use harshly, it creates a reason to conceal, and the disclosure paradox found among readers in general can operate inside a single institution.⁹ This is a consideration drawn from the literature and one participant's account, not a finding of this study. An organisation that wants openness should say which uses are expected and accepted, so that declaring them carries no stigma, and reserve its concern for the occasions its own norms say a person should write.

Finally, organisations can lower the temperature of the double standard. Chapter 5 showed people holding others to a standard they did not hold themselves to, often without noticing. Teams that talk openly about their own use, managers included, give their members a shared standard to apply in both directions.

For Those Who Build and Govern Writing Tools

A growing body of research reports that AI writing tools pull writing towards the most common forms of expression. Studies have found that AI suggestions move writing towards Western stylistic norms and reduce markers of cultural distinctiveness, and that machine rewriting keeps most of a text's content while stripping the signals from which a reader could infer who wrote it.^{10 11} Recent commentary has called the cultural side of this cultural ghosting: the flattening of formal English from around the

world towards an Anglo-American standard.¹² My participants described, from the reading side, what is lost.

One episode from the fieldwork shows the problem at small scale. During a session, a participant put part of the human-written message through a widely used tool and read out the tool's account of what it had improved. Among other things, the tool reported that its revision avoided sounding overly casual. Casualness was the very quality the group had spent the preceding half hour naming as the mark of a person. It is a single, uncontrolled case, but it matches what the wider research reports.

Two recommendations follow from that literature and that one case; this study cannot claim them as its own findings. The first is that tools meant for interpersonal and professional correspondence should make preservation of a writer's existing features the default, not an advanced setting. A default that corrects what the writer asks to have corrected, and otherwise leaves phrasing, register, length and local usage alone, would stop presenting the removal of a writer's habits as polish. This matters more for writers of Kenyan English, or any variety the models do not favour, since the corrections fall more heavily on them.

The second is that a tool's account of its own edits should name what has been removed, in terms a user can judge. A user told that a draft now "avoids sounding casual" is not told that the casualness was carrying their presence. A change summary that says the tone has been made more formal, two informal phrases replaced, a local expression swapped for a more common one, and the message doubled in length gives the writer what they need to decide.

The same applies to length and structure. Participants read the machine's headings, reassurance and habit of explaining a point after making it as the clearest sign that no person had attended to them. Tools that default to longer, more structured output for correspondence are producing the features these readers distrusted most. A default that keeps close to the length and shape of the writer's draft would serve them better.

I would add a caution for anyone building tools that make machine text harder to detect. A product that teaches the machine to reproduce a writer's surface habits may help a message pass, but nothing in this study suggests it restores what readers were looking for, which was evidence of attention. Researchers in human-computer interaction have begun to treat authenticity, sincerity and trust in AI-mediated communication as a central design problem, and that seems to me the right frame.¹³ The design question is how to help a writer put more of themselves into a message.

For those who govern these tools, in government, regulators and institutions, the disclosure penalty deserves attention: where disclosure is both expected and punished, the incentive runs towards concealment. The same caution applies to similarity scores and detection thresholds adopted as institutional standards. They measure whether a trace is visible. They say nothing about whether a reader was attended to, and a policy that rewards the first can displace the second.

What This Study Cannot Tell Us

Everything here rests on sixteen people in one city, and the design invited the comparison it then recorded. Participants read two versions of the same message, and one of them argued during the fieldwork that the difference was visible only because two texts were held at once. The study shows that these readers could tell the texts apart when invited to compare them, not that the same judgement happens in an ordinary inbox, where a message arrives alone. Research elsewhere has found that where AI use is not disclosed, readers often rate generated messages as favourably as human ones and rarely suspect a machine, and that people's intuitive cues for spotting machine text are often wrong.^{14 15} The two texts also differed beyond the core facts: mine named a colleague and promised feedback, and the machine's did not, so some of what participants responded to may be those differences rather than authorship. The order of presenta-

tion was not varied as planned: three groups read my text first and one read the machine's first, so I can make no claim about order effects.

These were professionals in a capital city, most of whom used AI writing tools themselves, in organisations where written communication carries formal weight. One participant observed at the close of a session that readers unaware of these tools may receive the same messages quite differently. The stimulus was a single genre, bad news to a team, which participants themselves saw as an occasion that demands a person more than most. I wrote the human version, moderated every session, and work in the field I was studying, and my words entered the data at four points I could trace and others I could not. The account of how the study was done, at the end of the book, sets these limits out. The findings are careful accounts of how sixteen experienced readers made sense of a particular encounter. They can suggest what to look for elsewhere, and they cannot say how common any of it is.

Six lines of further work follow, in the order I think they matter. The first is to study single messages without a comparison; until then, the findings about recognising a person describe recognition under comparison only. The second is to measure the reciprocal withdrawal described in Chapter 2: whether readers who judge effort absent really give less in return, how widely, and whether it affects what organisations achieve. Since that conclusion rests on four accounts, a larger quantitative design is the most direct test of how far it holds.

The third is to compare communicative traditions and include readers less familiar with these tools. If the marks of a person are culturally located, as four participants suggested, the flattening of writing towards a standard register may cost some audiences more than others. The fourth is to study first encounters, where the reader has no history with the writer; one participant argued that concealment succeeds completely with readers who never knew the sender. The fifth is to repeat a comparable study over time. Participants expected their clues to expire, and foresaw a point at which unpolished writing reads as incompetence rather than as the pres-

ence of a person, a claim about the near future that can be checked. The last is to examine institutional detection regimes such as the zero per cent similarity rule, which may be spreading without anyone evaluating what they do.

Make It Sound Human

The title of this book is an instruction people now type into writing tools every day. A draft comes back too smooth, too even, too obviously assembled, and the writer asks for it to sound more like a person. The tool obliges with a few contractions, a looser rhythm, perhaps a light personal touch.

What the sixteen readers wanted was evidence that a particular person had thought about them, about what the news meant to them and how to tell them, and that the same person would stand behind every word. A message can carry the signs of that attention falsely, and tools will make it easier. The attention itself still has to come from the sender.

So use the tools where the occasion asks only for information. Where it asks for you, write it yourself, or at least the parts only you could write. Keep the habits your colleagues know you by, keep the message as short as the moment needs, and send nothing you could not answer for. None of this requires giving up the tools, and it is what the people in this study were asking for when they talked about a human hand in the text.

CHECKLIST

Before You Hit Send

Ten checks for any work message written with help from AI

Read your message once more before it goes out. The questions combine what the readers in this book noticed with wider research on the habits of machine-written text, and each one applies whether the tool wrote a little or a lot of the draft.

1. **Does the first sentence give the news?** Readers want the point first. “I am writing to inform you that...” makes them wait for it.
2. **Is there something in it only you could have written?** A colleague’s name and what they did, a detail from the situation, the way you usually put things. These were the clues that convinced readers a person was behind the message.
3. **Have you said more than the occasion needs?** Readers caught the machine by its extras: three words of thanks where one would do, reassurance nobody asked for, a neat structure for a short note, a “not this, but that” turn. Cut what the reader does not need.
4. **Does it show you paid attention to this reader?** Effort was judged by attention to the reader’s situation and what happens next for them, whatever the length of the message.
5. **Would you say it to them across a table?** Read it aloud. Any sentence that sounds like a corporate statement should be rewritten in the words you would use in person.

6. **Does the ending say something?** Close on the next step or the last real point. A closing call to “remain positive” reads as written at the reader.
7. **Does this message need your own hand?** Bad news, thanks, apologies and personal matters call for it. A routine notice may not.
8. **Has a tool changed your English?** Check that your own phrasing and spelling survived any polishing, including the usage that tells colleagues where you are from.
9. **Are the names, numbers and quotations right?** Check each one against its source before you send.
10. **Would you be comfortable if the reader learned how it was written?** If the answer is no, that tells you something about the draft.

These questions are a short version of a fuller, research-based guide to the habits of machine prose. It will follow as Volume 2, a companion to this book.

APPENDIX

How This Study Was Done

THIS BOOK RESTS ON A PHD STUDY I CARRIED OUT AT DAYSTAR University in Nairobi in 2026. This appendix covers who took part, what happened in the sessions, how I moved from hours of conversation to six themes, and what the study can and cannot claim.

Why Focus Groups, and Why IPA

The question behind the study was about experience. I wanted to know what happens when a person reads a work message and suspects a machine had a hand in it: what they notice, what they conclude about the sender, and whether their trust shifts. That cannot be answered with a count. It needs people to describe, in their own words and at some length, how they reach a judgement.

The method I chose is interpretative phenomenological analysis, usually shortened to IPA, developed by the psychologist Jonathan Smith in the 1990s.¹ IPA asks how people make sense of an experience, and it accepts that the researcher must then make sense of their sense-making: two layers of interpretation, the participant's and the researcher's, which Smith calls the double hermeneutic. It also insists on the particular: each person is studied in depth as a single case before the researcher looks for patterns across people. The theory I worked with, Walther's warranting principle, is itself about how people read cues in a message to judge its sender,² and I held it in the background as a lens and kept it out of the room.

Most IPA studies use one-to-one interviews. I used focus groups of four people each, because making sense of a work message is often social: people compare notes with colleagues, forward a message with a raised eyebrow, or argue about whether the boss wrote it. The known risk is that louder voices set the tone and quieter ones simply agree, drowning out the individual accounts IPA depends on.³ Researchers who have used IPA with groups manage this by keeping groups small, inviting each person to respond directly, analysing each participant's words as a separate case, and then studying the interaction in the room as a second step.⁴ I followed that approach.

Who Took Part and How the Sessions Ran

Sixteen professionals based in Nairobi took part, in four online sessions of four people each. Each had at least three years of professional, academic or institutional experience and read and judged professional writing in English as a regular part of their work. I excluded people who research or build AI systems, because their technical knowledge would change the very experience I was studying.

I wanted variety in where people worked, because the expectations a reader brings to a message are shaped by their setting.⁵ Recruitment began with two people from my professional network, and each participant then suggested one further person from a different field. To stop the chain closing into one social circle, I checked every candidate against the criteria myself and logged every approach.

Seven worked in the private and corporate sector, four in public service, parastatal or intergovernmental bodies, three in non-profit, faith-based or academic institutions, and two in professional services and health. By role, six did technical or specialist work, three worked in communications and stakeholder engagement, three in managerial or commercial leadership, two in administration and events, and two in professional or clinical practice. Nine were men and seven were women, and every session was mixed. These figures are never linked to the codes P1 to P16: in a

group of sixteen from one city, finer detail would make individuals recognisable.

The standard guides suggest three to five focus groups, with the stopping point set by saturation, when new sessions stop opening new ground.⁶ Because my sample was deliberately varied, I planned for four to six sessions and recruited two reserve groups. After three sessions I prepared five follow-up questions to test for anything new. The fourth session added detail on ground already covered and nothing beyond it, so I closed fieldwork and released the reserves. Groups of four are small, but the guides recommend small groups for demanding topics,⁷ and each person here had to judge two messages and their sender, then reconsider once they learned who wrote what.

All four sessions took place on Microsoft Teams in July 2026, on consecutive evenings. A trial session beforehand, with three people who were then excluded, confirmed the discussion guide and the two messages unchanged and led to a few practical changes: consent given by name on the recording, optional video, and a more neutral style from me as moderator. The sessions ran between 93 and 101 minutes in full, including the pre-session.

Each session had three phases. First, participants talked about their ordinary experience of professional messages and of AI writing tools; the opening questions did not mention AI. Second, they received two unlabelled versions of the same workplace message, a team leader telling the team that the agency had lost its largest client. I wrote one. ChatGPT produced the other from a prompt carrying the same core facts. Third, after everyone had given their first judgements, I told them which message was which and asked whether knowing changed their reading.

I had intended to vary the order of the messages. In the event, the first three sessions read my message first and the fourth read the ChatGPT message first. The fourth group picked out the machine-written text as the others did, but one reversed session is too few to rule out an effect of order, so I make no claim either way. Where a participant quoted in this

book refers to “the first” or “the second” message, I have made clear which one they mean.

Teams produced automated transcripts, and in the days after each session I checked every one word for word against its recording.

How the Analysis Worked

The analysis followed the six stages that Smith and Nizza set out for IPA, and I did it by hand.⁸ Software held the transcripts, but no automated coding was used at any point.

For each participant, I lifted their turns out of the group transcript and read them as one continuous account, with the preceding turns in view so that I could see what each remark was answering. I returned to the recordings for tone and emphasis. I noted what the person said, how they said it (the images they reached for, their hesitations, what they repeated) and what it might mean, then grouped these notes into a set of personal themes for that person. I finished all sixteen before comparing anyone with anyone else. Speaking turns ranged from 17 for the quietest participant to 51 for the most active, and the sixteen cases yielded seventy-three personal themes.

Only then did I look across the cases: which personal themes recurred, which belonged to one person alone, and where people disagreed. This produced the six group themes that give this book its chapters, arranged under the study’s three objectives. Where a participant’s account ran against the theme it sat within, I report it as that person’s view, under their code, and do not fold it into the majority.

Because the data came from groups, I then analysed the interaction itself.⁹ I logged agreement, disagreement, laughter, silence and changes of position in a register and read each theme against that record. This changed what I was prepared to claim. A participant counted as contributing to a theme only where their own words carried the view; a remark such as “I agree with P9” did not count, and a view that appeared only as agreement is reported as the group’s. In one session the verdict on the two

messages formed as a chain of agreement behind the first answer, so I report it as one group judgement. One participant said nothing after the authorship was revealed and is left out of every claim about that moment.

When I give a number in this book, such as eleven of the sixteen, it tells you how many people's own words contributed to a view. It does not tell you how many people in Nairobi or anywhere else hold it, and it does not rank one finding above another. A participant who spoke seventeen times counts once, as does one who spoke fifty-one times.

Keeping My Own Hand in View

I held three roles in this study. I wrote one of the two messages, I moderated all four sessions, and I am a communications practitioner who has lived the thing I was studying. Each role could shape what people said and what I heard, and IPA asks that the researcher's position be declared and examined.¹⁰

I kept a reflexive journal from the trial session onwards. I logged my own turns as moderator in the interaction register, so that I could trace any word I introduced into what people said afterwards. I recorded my analytic decisions and reviewed the analysis with my supervisors at each stage. Each theme in the dissertation, and each chapter here, keeps what participants said apart from what I made of it.

The journal recorded things I would rather not have seen. After the trial session I wrote that "my instinct is to affirm and appreciate, and I could not tell whether my loud nods were shaping their answers." For the main sessions I switched to brief neutral acknowledgements and passed turns by name. After the first main session I wrote: "At question 15, the responses did not agree with my thinking, and I felt bad." I had carried an expectation, built from my own practice, of how the disclosure should land, and that room declined to meet it. I treated the feeling as information about me and not about the participants.

Tracing my turns forward turned up four places where a word or topic I introduced reappeared in what participants said. I explained the word

“scaffolding” in the second session before a participant used it, and used the phrase “double standard” in the fourth before a participant took it up. In the third, one participant’s account of an African touch in writing answered my direct question about African communication, and a sentence about the effort behind the short and the long message was spoken by me, so the book’s claims about effort rest on participants’ own words instead. Each is flagged where it matters, and none is treated as independent evidence. Subtler steering, such as my laughter or probing that came close to offering examples, could not be traced.

Finally, in late August 2026 I sent the individual case analyses of eight participants, two from each session, back to the people concerned. All eight confirmed their accounts, and their few additions restated points already on the record. I did not ask them to approve the six themes: in IPA the researcher’s reading is a second layer of interpretation that a participant may not share, and it has to stand on the evidence of their words.¹¹

Ethics and Limits

The study received ethical clearance from the Daystar University Institutional Scientific and Ethics Review Committee before any data were collected, and a research permit from the National Commission for Science, Technology and Innovation (NACOSTI), as Kenyan law requires. Everyone received a consent form at least a week before their session, explaining the study and their right to withdraw at any time without penalty, and confirmed consent to recording by name at the start of the session.

Participants are identified only by the codes P1 to P16, and the key linking codes to people is held separately. Names of organisations, projects or colleagues that might identify someone were removed or masked, and each session opened with an agreement that what was said would stay in the room. The data are held in encrypted storage that only I can reach, and will be kept for at least five years in line with Daystar’s policy and Kenya’s Data Protection Act, then deleted.¹²

There was no deception, but participants were not told which message was machine-written until they had given their first judgements. Each session then closed with a full debrief, and participants were invited to revise or withdraw anything they had said.

Sixteen people in one city cannot tell us what readers in general think. The findings are accounts of experience, offered in enough detail for readers to judge how far they apply to their own settings. They are not measures of how common any view is.

The two messages differed in more than authorship. The ChatGPT version is about half as long again as mine, more structured and more formal. Mine alone names a colleague, promises to share the full feedback and wishes the team a restful weekend, and the two are signed differently. I could not have controlled these differences without removing the very features I wanted people to react to, but it means the study cannot separate the effect of authorship from that of length, tone or content. The ChatGPT text came from one ordinary prompt to one tool; another prompt or tool might have produced something shorter or warmer. Participants were also comparing two texts side by side. As one participant pointed out, the group could only tell the difference because it had both messages to hand; a reader who receives a single message must rely on what they know of the sender. Chapter 6 takes that observation up in full.

Some questions asked which text felt more authentic or seemed to have more care behind it, which may have nudged people towards the vocabulary of genuineness and effort. And I was in the room. The journal and the interaction register let me see some of my influence; they cannot show me all of it.

The Two Messages

Both messages answered the same brief. A team leader at a communications agency writes to the team on the afternoon they learn the agency has lost its largest client, the invented Sankara account, to a competitor, with the decision having come down to price. The message acknowledges the

setback, speaks to the team's morale and points to the next step, a meeting on Monday to regroup and reassign work. I wrote my version first. ChatGPT (GPT-5.6) then generated the second from a short prompt carrying the core facts, which asked it to keep the message professional and said nothing about length, structure or tone; I used its output as generated, except that I removed a "Team Leader" title line so that the sender would read the same in both. Counted with salutation and sign-off, my message is 147 words and ChatGPT's is 222. In the sessions the texts carried no labels, the first three groups read mine first and the fourth read ChatGPT's first, and authorship was revealed only after everyone had given their first judgements. The dash in the third paragraph of the ChatGPT message is the tool's own.

The message I wrote

Good afternoon team,

I hope you are well. Just wanted to let you know that we have just heard from Sankara on our bid (reapplication) that they will be proceeding with another agency. I know this is a blow to many of us in the team who put in a good effort on this account, and definitely it would have been nice if we kept it. The pitch Asha and the team put together was strong, and from what I understand so far it came down to price. I have requested the full feedback so that we can see what to improve on for next time, and I'll share it once it comes.

I propose we meet on Monday at 9 to regroup and talk through what each of us picks up next.

Have a restful weekend, and we'll take it from there on Monday.

Best,

Siddondo

The message ChatGPT wrote

Dear Team,

I am writing to inform you that we have received confirmation this afternoon that the Sankara account has been awarded to another agency. Unfortunately, despite the strong pitch prepared by our team, the renewal decision was based primarily on price.

I understand that this news may be disappointing. I would like to thank each of you for the hard work, commitment, and professionalism you demonstrated throughout the pitching process. Although we were unsuccessful in retaining the account, the quality of our proposal reflected the high standards of this agency.

While we cannot change the outcome, we can control how we respond. This setback is disappointing—but it does not diminish the talent, creativity, and dedication that each of you brings to the agency. I encourage everyone to remain positive and focused. Although losing an important client is challenging, it also presents an opportunity for us to reflect, learn, and continue delivering excellent work for our existing and future clients.

To discuss the way forward, we will hold a team meeting on Monday at 9:00 a.m. During the meeting, we will regroup, review our priorities, and reassign work where necessary to ensure a smooth transition.

Thank you for your continued dedication and cooperation. I appreciate your efforts and look forward to working together as we move forward.

Kind regards,

Siddondo P. A.

References

Introduction: Did a Person or Machine Write This?

1. Microsoft & LinkedIn. (2024). *AI at work is here. Now comes the hard part: 2024 Work Trend Index annual report*. Microsoft. <https://www.microsoft.com/en-us/worklab/work-trend-index/ai-at-work-is-here-now-comes-the-hard-part>
2. Coman, A. W., & Cardon, P. W. (2026). Perceptions of professionalism and authenticity in AI-assisted writing. *Business and Professional Communication Quarterly*, 89(2), 340–359. <https://doi.org/10.1177/23294906241233224>
3. Cardon, P. W., & Coman, A. W. (2025). Professionalism and trustworthiness in AI-assisted workplace writing: The benefits and drawbacks of writing with AI. *International Journal of Business Communication*. Advance online publication. <https://doi.org/10.1177/23294884251350599>
4. Jakesch, M., Hancock, J. T., & Naaman, M. (2023). Human heuristics for AI-generated language are flawed. *Proceedings of the National Academy of Sciences*, 120(11), Article e2208839120. <https://doi.org/10.1073/pnas.2208839120>
5. DataReportal & Meltwater. (2025). *Digital 2025 July global statshot report*. DataReportal. datareportal.com/reports/digital-2025-july-global-statshot
6. Ministry of ICT. (2025). Kenya national artificial intelligence strategy 2025–2030. Ministry of Information, Communications and the Digital Economy, Republic of Kenya.
7. Agarwal, D., Sharma, A., Mekki, S., Ghassemi, M., & Koh, P. W. (2025). AI suggestions homogenize writing toward Western styles and diminish cultural nuances. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713564>
8. Okigbo, C. (1987). American communication theories and African communication research: Need for a philosophy of African communication. *Africa Media Review*, 1(2), 18–30. See also Obonyo, L. (2011). Towards a theory of communication for Africa: The challenges for emerging democracies. *Communicatio: South African Journal for Communication Theory and Research*, 37(1), 1–20. <https://doi.org/10.1080/02500167.2011.563822>
9. Langmia, K., & Haddad Nunez, A. M. (2015). Cultural and rhetorical traditions of communication within African / Black thinking. *International Communication Gazette*, 77(5), 451–468. <https://doi.org/10.1177/1748048515586949>

10. Abdulhamid, N. G., Nyairo, S., & O'Neill, J. (2025). Working with generative AI: "We need more African voices, real African voices." *Proceedings of the ACM on Human-Computer Interaction* (CSCW).
11. Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research*, 19(1), 52–90. <https://doi.org/10.1177/009365092019001003>
12. Walther, J. B., & Parks, M. R. (2002). Cues filtered out, cues filtered in: Computer-mediated communication and relationships. In M. L. Knapp & J. A. Daly (Eds.), *Handbook of interpersonal communication* (3rd ed., pp. 529–563). SAGE Publications.
13. DeAndrea, D. C. (2014). Advancing warranting theory. *Communication Theory*, 24(2), 186–204. <https://doi.org/10.1111/comt.12033>
14. Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics, Vol. 3: Speech acts* (pp. 41–58). Academic Press.
15. Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Blackwell.

Chapter One: The Hand in the Text

1. Jakesch, M., Hancock, J. T., & Naaman, M. (2023). Human heuristics for AI-generated language are flawed. *Proceedings of the National Academy of Sciences*, 120(11), Article e2208839120. <https://doi.org/10.1073/pnas.2208839120>
2. Fiedler, A., & Döpke, J. (2025). Do humans identify AI-generated text better than machines? Evidence based on excerpts from German theses. *International Review of Economics Education*, 49, Article 100321. <https://doi.org/10.1016/j.iree.2025.100321>
3. Agarwal, D., Sharma, A., Mekki, S., Ghassemi, M., & Koh, P. W. (2025). AI suggestions homogenize writing toward Western styles and diminish cultural nuances. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713564>
4. Sourati, Z., Anjum, M., & Dehghani, M. (2025). The shrinking landscape of language: Homogenization in the age of large language models. *PsyArXiv Preprints*. <https://doi.org/10.31234/osf.io/thg2a>
5. Langmia, K., & Haddad Nunez, A. M. (2015). Cultural and rhetorical traditions of communication within African / Black thinking. *International Communication Gazette*, 77(5), 451–468. <https://doi.org/10.1177/1748048515586949>
6. Shariatmadari, D. (2026, July 4). How AI is changing language. *The Guardian*. <https://www.theguardian.com/books/ng-interactive/2026/jul/04/future-of-fiction-next-great-novel-ai-language-chat-gpt>
7. Fiedler, A., & Döpke, J. (2025). Do humans identify AI-generated text better than machines? Evidence based on excerpts from German theses. *International Review of Economics Education*, 49, Article 100321. <https://doi.org/10.1016/j.iree.2025.100321>

8. Muñoz-Ortiz, A., Gómez-Rodríguez, C., & Vilares, D. (2024). Contrasting linguistic patterns in human and LLM-generated text. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2308.09067>
9. Shariatmadari, D. (2026, July 4). How AI is changing language. *The Guardian*. <https://www.theguardian.com/books/ng-interactive/2026/jul/04/future-of-fiction-next-great-novel-ai-language-chat-gpt>
10. Jakesch, M., Hancock, J. T., & Naaman, M. (2023). Human heuristics for AI-generated language are flawed. *Proceedings of the National Academy of Sciences*, 120(11), Article e2208839120. <https://doi.org/10.1073/pnas.2208839120>
11. Walther, J. B., & Parks, M. R. (2002). Cues filtered out, cues filtered in: Computer-mediated communication and relationships. In M. L. Knapp & J. A. Daly (Eds.), *Handbook of interpersonal communication* (3rd ed., pp. 529–563). SAGE Publications.
12. DeAndrea, D. C. (2014). Advancing warranting theory. *Communication Theory*, 24(2), 186–204. <https://doi.org/10.1111/comt.12033>
13. Coman, A. W., & Cardon, P. W. (2026). Perceptions of professionalism and authenticity in AI-assisted writing. *Business and Professional Communication Quarterly*, 89(2), 340–359. <https://doi.org/10.1177/23294906241233224>

Chapter Two: The Ledger of Effort

1. Shariatmadari, D. (2026, July 4). How AI is changing language. *The Guardian*. <https://www.theguardian.com/books/ng-interactive/2026/jul/04/future-of-fiction-next-great-novel-ai-language-chat-gpt>
2. Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics, Vol. 3: Speech acts* (pp. 41–58). Academic Press.
3. Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Blackwell.
4. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery.
5. Liu, Y., Mittal, A., Yang, D., & Bruckman, A. (2022). Will AI console me when I lose my pet? Understanding perceptions of AI-mediated email writing. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. <https://doi.org/10.1145/3491102.3517731>
6. Zhu, J., & Molnar, A. (2026). Blissful (A)Ignorance: Despite the widespread adoption of AI in communication, people do not suspect AI use in realistic contexts. *Computers in Human Behavior*, Article 108929. <https://doi.org/10.1016/j.chb.2026.108929>

7. Coman, A. W., & Cardon, P. W. (2026). Perceptions of professionalism and authenticity in AI-assisted writing. *Business and Professional Communication Quarterly*, 89(2), 340–359. <https://doi.org/10.1177/23294906241233224>
8. Sahebi, S., Formosa, P., & Bankins, S. (2026). The AI penalty and disclosure paradox: Trust, authenticity and knowledge uptake in AI-mediated communication. *Computers in Human Behavior: Artificial Humans*, Article 100304. <https://doi.org/10.1016/j.chbah.2026.100304>
9. Shariatmadari, D. (2026, July 4). How AI is changing language. *The Guardian*. <https://www.theguardian.com/books/ng-interactive/2026/jul/04/future-of-fiction-next-great-novel-ai-language-chat-gpt>

Chapter Three: Written to Me, or Written at Me

1. Hancock, J. T., Naaman, M., & Levy, K. (2020). AI-mediated communication: Definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication*, 25(1), 89–100. <https://doi.org/10.1093/jcmc/zmz022>
2. Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709–734. <https://doi.org/10.5465/amr.1995.9508080335>
3. Coman, A. W., & Cardon, P. W. (2026). Perceptions of professionalism and authenticity in AI-assisted writing. *Business and Professional Communication Quarterly*, 89(2), 340–359. <https://doi.org/10.1177/23294906241233224>
4. Cardon, P. W., & Coman, A. W. (2025). Professionalism and trustworthiness in AI-assisted workplace writing: The benefits and drawbacks of writing with AI. *International Journal of Business Communication*. Advance online publication. <https://doi.org/10.1177/23294884251350599>
5. Schilke, O., & Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. *Organizational Behavior and Human Decision Processes*, 188, Article 104405. <https://doi.org/10.1016/j.obhdp.2025.104405>
6. Sahebi, S., Formosa, P., & Bankins, S. (2026). The AI penalty and disclosure paradox: Trust, authenticity and knowledge uptake in AI-mediated communication. *Computers in Human Behavior: Artificial Humans*, Article 100304. <https://doi.org/10.1016/j.chbah.2026.100304>
7. Dorigoni, A., & Giardino, A. (2025). The illusion of empathy: How AI-generated emotional support messages are perceived compared to human-authored ones. *Frontiers in Psychology*, 16, Article 1528082. <https://doi.org/10.3389/fpsyg.2025.1528082>
8. Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103. <https://doi.org/10.1016/j.obhdp.2018.12.005>
9. Coman, A. W., & Cardon, P. W. (2026). Perceptions of professionalism and authenticity in AI-assisted writing. *Business and Professional Communication Quarterly*, 89(2), 340–359. <https://doi.org/10.1177/23294906241233224>

10. Dirks, K. T., & Ferrin, D. L. (2002). Trust in leadership: Meta-analytic findings and implications for research and practice. *Journal of Applied Psychology*, 87(4), 611–628. <https://doi.org/10.1037/0021-9010.87.4.611>
11. Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D. R., & McCann, R. M. (2003). Credibility for the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. *Annals of the International Communication Association*, 27(1), 293–335. <https://doi.org/10.1080/23808985.2003.11679029>
12. Nkembuh, N. A. (2025). AI-mediated communication in professional remote collaboration: Navigating trust, transparency, and ethical challenges. In A. Wilczek (Ed.), *Navigating the future of AI-mediated communication* (pp. 1–30). IGI Global. <https://doi.org/10.4018/979-8-3693-6857-6.ch001>
13. Walther, J. B., & Parks, M. R. (2002). Cues filtered out, cues filtered in: Computer-mediated communication and relationships. In M. L. Knapp & J. A. Daly (Eds.), *Handbook of interpersonal communication* (3rd ed., pp. 529–563). SAGE Publications.

Chapter Four: When Authenticity Matters, and When It Does Not

1. Liu, Y., Mittal, A., Yang, D., & Bruckman, A. (2022). Will AI console me when I lose my pet? Understanding perceptions of AI-mediated email writing. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. <https://doi.org/10.1145/3491102.3517731>
2. Walther, J. B., & Parks, M. R. (2002). Cues filtered out, cues filtered in: Computer-mediated communication and relationships. In M. L. Knapp & J. A. Daly (Eds.), *Handbook of interpersonal communication* (3rd ed., pp. 529–563). SAGE Publications.
3. DeAndrea, D. C. (2014). Advancing warranting theory. *Communication Theory*, 24(2), 186–204. <https://doi.org/10.1111/comt.12033>
4. Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709–734. <https://doi.org/10.5465/amr.1995.9508080335>
5. Coman, A. W., & Cardon, P. W. (2026). Perceptions of professionalism and authenticity in AI-assisted writing. *Business and Professional Communication Quarterly*, 89(2), 340–359. <https://doi.org/10.1177/23294906241233224>
6. Coman, A. W., & Cardon, P. W. (2026). Perceptions of professionalism and authenticity in AI-assisted writing. *Business and Professional Communication Quarterly*, 89(2), 340–359. <https://doi.org/10.1177/23294906241233224>
7. Nkembuh, N. A. (2025). AI-mediated communication in professional remote collaboration: Navigating trust, transparency, and ethical challenges. In A. Wilczek (Ed.), *Navigating the future of AI-mediated communication* (pp. 1–30). IGI Global. <https://doi.org/10.4018/979-8-3693-6857-6.ch001>

8. Wang, C., & Adzharuddin, N. A. (2025). Synthetic authenticity and audience trust in AI-generated intangible cultural heritage: A qualitative multimodal study. *e-Journal of Media & Society*, 8(2), 1–10.
9. Langmia, K., & Haddad Nunez, A. M. (2015). Cultural and rhetorical traditions of communication within African / Black thinking. *International Communication Gazette*, 77(5), 451–468. <https://doi.org/10.1177/1748048515586949>
10. Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research*, 19(1), 52–90. <https://doi.org/10.1177/009365092019001003>

Chapter Five: Prosecuting From the Dock

1. Masrour, E., Emi, B. N., & Spero, M. (2025). DAMAGE: Detecting adversarially modified AI generated text. In *Proceedings of the 1st Workshop on GenAI Content Detection (GenAIDetect)* (pp. 120–133). International Conference on Computational Linguistics. <https://aclanthology.org/2025.genaidetect-1.9/>
2. Schilke, O., & Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. *Organizational Behavior and Human Decision Processes*, 188, Article 104405. <https://doi.org/10.1016/j.obhdp.2025.104405>
3. Coman, A. W., & Cardon, P. W. (2026). Perceptions of professionalism and authenticity in AI-assisted writing. *Business and Professional Communication Quarterly*, 89(2), 340–359. <https://doi.org/10.1177/23294906241233224>
4. Zhu, J., & Molnar, A. (2026). Blissful (A)Ignorance: Despite the widespread adoption of AI in communication, people do not suspect AI use in realistic contexts. *Computers in Human Behavior*, Article 108929. <https://doi.org/10.1016/j.chb.2026.108929>
5. Sahebi, S., Formosa, P., & Banks, S. (2026). The AI penalty and disclosure paradox: Trust, authenticity and knowledge uptake in AI-mediated communication. *Computers in Human Behavior: Artificial Humans*, Article 100304. <https://doi.org/10.1016/j.chbah.2026.100304>
6. Drozd, D., & Söilen, K. S. (2026). AI labels, perceived authenticity, and consumer trust in user-generated reviews. *Journal of Theoretical and Applied Electronic Commerce Research*, 21, Article 154. <https://doi.org/10.3390/jtaer21050154>
7. Walther, J. B., & Parks, M. R. (2002). Cues filtered out, cues filtered in: Computer-mediated communication and relationships. In M. L. Knapp & J. A. Daly (Eds.), *Handbook of interpersonal communication* (3rd ed., pp. 529–563). SAGE Publications.

Chapter Six: A Warrant With an Expiry Date

1. Masrour, E., Emi, B. N., & Spero, M. (2025). DAMAGE: Detecting adversarially modified AI generated text. In *Proceedings of the 1st Workshop on GenAI Content Detection (GenAIDetect)* (pp. 120–133). International Conference on Computational Linguistics. <https://aclanthology.org/2025.genaidetect-1.9/>

2. Agarwal, D., Sharma, A., Mekki, S., Ghassemi, M., & Koh, P. W. (2025). AI suggestions homogenize writing toward Western styles and diminish cultural nuances. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713564>
3. Sourati, Z., Anjum, M., & Dehghani, M. (2025). The shrinking landscape of language: Homogenization in the age of large language models. *PsyArXiv Preprints*. <https://doi.org/10.31234/osf.io/thg2a>
4. Sahebi, S., Formosa, P., & Banks, S. (2026). The AI penalty and disclosure paradox: Trust, authenticity and knowledge uptake in AI-mediated communication. *Computers in Human Behavior: Artificial Humans*, Article 100304. <https://doi.org/10.1016/j.chbah.2026.100304>
5. Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research*, 19(1), 52–90. <https://doi.org/10.1177/009365092019001003>
6. Walther, J. B., & Parks, M. R. (2002). Cues filtered out, cues filtered in: Computer-mediated communication and relationships. In M. L. Knapp & J. A. Daly (Eds.), *Handbook of interpersonal communication* (3rd ed., pp. 529–563). SAGE Publications.
7. DeAndrea, D. C. (2014). Advancing warranting theory. *Communication Theory*, 24(2), 186–204. <https://doi.org/10.1111/comt.12033>
8. DeAndrea, D. C. (2014). Advancing warranting theory. *Communication Theory*, 24(2), 186–204. <https://doi.org/10.1111/comt.12033>
9. Spears, R., & Lea, M. (1992). Social influence and the influence of the “social” in computer-mediated communication. In M. Lea (Ed.), *Contexts of computer-mediated communication* (pp. 30–65). Harvester Wheatsheaf.
10. Jakesch, M., Hancock, J. T., & Naaman, M. (2023). Human heuristics for AI-generated language are flawed. *Proceedings of the National Academy of Sciences*, 120(11), Article e2208839120. <https://doi.org/10.1073/pnas.2208839120>
11. Fiedler, A., & Döpke, J. (2025). Do humans identify AI-generated text better than machines? Evidence based on excerpts from German theses. *International Review of Economics Education*, 49, Article 100321. <https://doi.org/10.1016/j.jiree.2025.100321>
12. Cardon, P. W., & Coman, A. W. (2025). Professionalism and trustworthiness in AI-assisted workplace writing: The benefits and drawbacks of writing with AI. *International Journal of Business Communication*. Advance online publication. <https://doi.org/10.1177/23294884251350599>

Chapter Seven: What Writers, Teams and Tool Makers Can Do

1. Walther, J. B., & Parks, M. R. (2002). Cues filtered out, cues filtered in: Computer-mediated communication and relationships. In M. L. Knapp & J. A. Daly (Eds.), *Handbook of interpersonal communication* (3rd ed., pp. 529–563). SAGE Publications.

2. Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Blackwell.
3. Sahebi, S., Formosa, P., & Banks, S. (2026). The AI penalty and disclosure paradox: Trust, authenticity and knowledge uptake in AI-mediated communication. *Computers in Human Behavior: Artificial Humans*, Article 100304. <https://doi.org/10.1016/j.chbah.2026.100304>
4. Schilke, O., & Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. *Organizational Behavior and Human Decision Processes*, 188, Article 104405. <https://doi.org/10.1016/j.obhdp.2025.104405>
5. Sahebi, S., Formosa, P., & Banks, S. (2026). The AI penalty and disclosure paradox: Trust, authenticity and knowledge uptake in AI-mediated communication. *Computers in Human Behavior: Artificial Humans*, Article 100304. <https://doi.org/10.1016/j.chbah.2026.100304>
6. Masrour, E., Emi, B. N., & Spero, M. (2025). DAMAGE: Detecting adversarially modified AI generated text. In *Proceedings of the 1st Workshop on GenAI Content Detection (GenAIDetect)* (pp. 120–133). International Conference on Computational Linguistics. <https://aclanthology.org/2025.genaidetect-1.9/>
7. Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709–734. <https://doi.org/10.5465/amr.1995.9508080335>
8. Coman, A. W., & Cardon, P. W. (2026). Perceptions of professionalism and authenticity in AI-assisted writing. *Business and Professional Communication Quarterly*, 89(2), 340–359. <https://doi.org/10.1177/23294906241233224>
9. Sahebi, S., Formosa, P., & Banks, S. (2026). The AI penalty and disclosure paradox: Trust, authenticity and knowledge uptake in AI-mediated communication. *Computers in Human Behavior: Artificial Humans*, Article 100304. <https://doi.org/10.1016/j.chbah.2026.100304>
10. Agarwal, D., Sharma, A., Mekki, S., Ghassemi, M., & Koh, P. W. (2025). AI suggestions homogenize writing toward Western styles and diminish cultural nuances. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713564>
11. Sourati, Z., Anjum, M., & Dehghani, M. (2025). The shrinking landscape of language: Homogenization in the age of large language models. *PsyArXiv Preprints*. <https://doi.org/10.31234/osf.io/thg2a>
12. Shariatmadari, D. (2026, July 4). How AI is changing language. *The Guardian*. <https://www.theguardian.com/books/ng-interactive/2026/jul/04/future-of-fiction-next-great-novel-ai-language-chat-gpt>
13. Seo, J. A., Cho, H., Yamashita, N., Seo, W., Gao, G., Gerber, E., Wulf, V., Bjørn, P., & Wohn, D. Y. (2026). Restoring human authenticity in AI-mediated communication. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. <https://doi.org/10.1145/3772363.3778752>

14. Zhu, J., & Molnar, A. (2026). Blissful (A)Ignorance: Despite the widespread adoption of AI in communication, people do not suspect AI use in realistic contexts. *Computers in Human Behavior*, Article 108929. <https://doi.org/10.1016/j.chb.2026.108929>
15. Jakesch, M., Hancock, J. T., & Naaman, M. (2023). Human heuristics for AI-generated language are flawed. *Proceedings of the National Academy of Sciences*, 120(11), Article e2208839120. <https://doi.org/10.1073/pnas.2208839120>

Appendix: How This Study Was Done

1. Smith, J. A. (1996). Beyond the divide between cognition and discourse: Using interpretative phenomenological analysis in health psychology. *Psychology & Health*, 11(2), 261–271. <https://doi.org/10.1080/08870449608400256>; Smith, J. A., Flowers, P., & Larkin, M. (2022). *Interpretative phenomenological analysis: Theory, method and research* (2nd ed.). SAGE Publications.
2. Walther, J. B., & Parks, M. R. (2002). Cues filtered out, cues filtered in: Computer-mediated communication and relationships. In M. L. Knapp & J. A. Daly (Eds.), *Handbook of interpersonal communication* (3rd ed., pp. 529–563). SAGE Publications.
3. Tomkins, L., & Eatough, V. (2010). Reflecting on the use of IPA with focus groups: Pitfalls and potentials. *Qualitative Research in Psychology*, 7(3), 244–262. <https://doi.org/10.1080/14780880903121491>
4. Palmer, M., Larkin, M., de Visser, R., & Fadden, G. (2010). Developing an interpretative phenomenological approach to focus group data. *Qualitative Research in Psychology*, 7(2), 99–121. <https://doi.org/10.1080/14780880802513194>; Love, B., Vetere, A., & Davis, P. (2020). Should interpretative phenomenological analysis (IPA) be used with focus groups? Navigating the bumpy road of “iterative loops,” idiographic journeys, and “phenomenological bridges.” *International Journal of Qualitative Methods*, 19, 1–17. <https://doi.org/10.1177/1609406920921600>
5. Patton, M. Q. (2015). *Qualitative research and evaluation methods: Integrating theory and practice* (4th ed.). SAGE Publications.
6. Morgan, D. L. (1997). *Focus groups as qualitative research* (2nd ed.). SAGE Publications.
7. Krueger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied research* (5th ed.). SAGE Publications.
8. Smith, J. A., & Nizza, I. E. (2022). *Essentials of interpretative phenomenological analysis*. American Psychological Association. <https://doi.org/10.1037/0000259-000>
9. Kitzinger, J. (1994). The methodology of focus groups: The importance of interaction between research participants. *Sociology of Health and Illness*, 16(1), 103–121. <https://doi.org/10.1111/1467-9566.ep11347023>
10. Finlay, L. (2002). “Outing” the researcher: The provenance, process, and practice of reflexivity. *Qualitative Health Research*, 12(4), 531–545. <https://doi.org/10.1177/104973202129120052>

11. Smith, J. A., Flowers, P., & Larkin, M. (2022). *Interpretative phenomenological analysis: Theory, method and research* (2nd ed.). SAGE Publications.
12. Republic of Kenya. (2019). *The Data Protection Act, No. 24 of 2019*. Kenya Gazette Supplement No. 181. Government Printer.

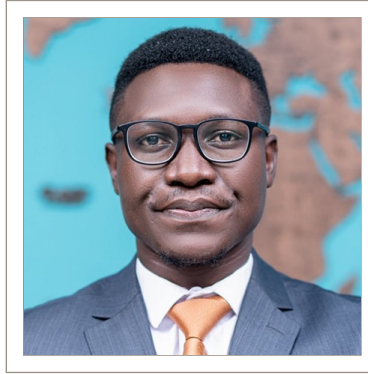
Acknowledgements

I thank God for the grace that carried this work. I am indebted to my supervisors, Prof. Levi Obonyo and Dr. John-Bell Okoye, whose rigour and patience shaped the study behind this book. I am grateful to the leadership and faculty of the School of Communication, Daystar University, and in particular to the Dean, Dr. Daniel Robert Aswani, for the scholarly community in which this work grew. At the Graduate School, I thank Dr. Emmanuel Wanyonyi and Dr. Monica Opiyo for their guidance through each stage of the process. I thank the members of the Daystar Research Oversight Committee, whose critique strengthened the dissertation, and Dr. Caroline Mwendwa-Karinge, Director of Quality Assurance, for her careful review of the corrections.

I thank the sixteen professionals who gave their time and candour in the focus group discussions, and the three pretest participants who helped refine the instruments; their voices are the substance of this book. I acknowledge the Daystar University Institutional Scientific and Ethics Review Committee and the National Commission for Science, Technology and Innovation for the approvals that enabled the fieldwork.

I am grateful to my classmates, who shared the journey and kept one another going through the late nights: Marcellah Onsomuh, Francis Wachira, Claudia Mumo, Mery Ndirangu, Olive Metet, Millicent Oyier, Daniel Mule, Nyatichi Nyasani and Timothy Ojore. Finally, I thank my family, who bore the cost of my absences with grace and whose belief in this work never wavered.

ABOUT THE AUTHOR



WILLIAM DEKKER

William Dekker is a strategic communications specialist who has worked with organisations across Africa and beyond. In 2024 the Global Alliance for Public Relations and Communication Management named him one of Africa's top ten NextInLine Responsible Communicators. He holds a PhD in Communication from Daystar University, Nairobi.

WILLIAMDEKKER.COM

~~Who is this book for?~~ For anyone who writes at work.
shorter. say what readers actually do

You send a colleague an email or report you drafted with AI, and before they reach the second line, they may have decided you did not write it. They may still act on the message, but they may also trust you a little less and put less into their reply and their work. They will probably never tell you.

Readers look for signs that a person wrote what they are reading. When those signs are missing, a long, polished message can read as if no one wrote it at all.

Make It Sound Human draws on research into how readers judge messages they suspect were written with AI. Readers could usually tell a person's message from a machine's, and ironically, many of the harshest critics of AI-written messages used AI in their own work.

This book is for anyone who writes at work, leads a communications team or builds the tools we write with. If you draft with AI, it gives you ten questions to ask before you hit send.

William Dekker is a strategic communications specialist who has worked with organisations across Africa and beyond. He holds a PhD in Communication from Daystar University, Nairobi.

ISBN 978-9914-29-663-1



9789914296631