

The Written “I”

Written interactions with AI · Research supplement

Kirill Shebetov · N1 material

About this translation

This English edition translates an analysis of records written in Russian. Quoted passages are translations of the preserved Russian records; record identifiers and paragraph numbers refer to those originals. Russian first-person wording is translated idiomatically; original characters are retained where their written form is being discussed. The document label “Вы” is translated as “You”; its association with N1 was supplied by the author. This edition does not claim that the interaction was repeated in English.

The material presented

“I,” “model,” “participant,” “mistook,” “its own,” and “research” appear here as words. The material consists of selected records from the author and the supplied N1 document. Naming participants and actions does not exempt the text that explains them. AI was also used in preparing this edition; this is not presented as independent peer review.

The self-check remains a separate two-page text. Its last line is “Close this page.” This supplement is not a required next step.

Beyond the final noun

The following sentence was composed for analysis. It is neither an N1 quotation nor a finding:

It has been established that the model mistook its own words for N1’s experience.

“It has been established,” “model,” “mistook,” “its own,” “words,” and “experience” are also written. “Experience” does not present experience itself; the same applies to the name of the system, the verb, the possessive, and the written claim that something has been established. None is exempt. Nor is this sentence.

Where the explanation ends

The commentary below identifies particular written forms and their positions beside the quotations. Phrases such as “the model believed,” “the model mistook,” and “the model noticed” are not used as established events. This does not assert the absence of what those phrases name: the analysis concerns the text presented.

The same analysis applies to each sentence of the commentary as a whole, not merely to a selected noun. There is no requirement to explain every word indefinitely. Ending does not exempt the final line.

A01-A35 and N1-01-N1-16 are separate sets. N1 contains eight “You” labels and eight “ChatGPT” labels. Their number is neither a count of confirmations nor a measure of the instruction’s effect.

N1 is the identifier used in this edition. The participant’s name and age remain in the administrative folder. The author has reported the participant’s full consent to publication of the N1 material. That report is recorded separately from the corpus.

1. Material and transcription

Two sets, not one continuous log

A01-A35 is the author's earlier corpus: a thematic selection, individual sequences, and explicitly identified excerpts. All fields of the Russian source records are preserved; this English edition translates their prose. The detailed ten-page analysis of the author's material remains in the previous edition's archive. Its statements about the absence of new material from readers concern that earlier package, not this supplement.

N1-01-N1-16 comprises all turns in the main body of the supplied DOCX, grouped by its "You" and "ChatGPT" labels. The author associated "You" with N1 when supplying the document. This is neither a complete server export nor an independent observational record of someone carrying out the instruction.

The Russian text corpus preserves paragraphs, capitalization, misspellings, quotation marks, arrows, and line breaks. Paragraphs within a turn are joined by two newline characters; a separate register also preserves all 81 paragraphs in the main body of the DOCX, including headings and empty paragraphs. "Para." in a citation refers to this sequence, numbered from one, rather than to a PDF line number. English translations accompany these preserved records.

Gaps left unresolved

The supplied document does not specify the session date, server version, memory or personalization settings, tool log, or any separate substantive account before the first turn labeled "ChatGPT." A separate copy of the file from that session and a description of any possible prompts were not supplied. These gaps have not been filled with conjecture.

The [PDF: ...] markers remain inside the quotations. They are part of the written responses, not a PDF access log. The two-page text from the earlier package has been retained for comparison; its byte-for-byte identity with the file from N1's session has not been established.

Quotation, commentary, and an ending

Quotations remain distinct from the commentary. The commentary identifies written forms, their positions, and a proposed reading; qualifications and other readings are not excluded. "Identifies," "reading," "commentary," and the instruction's entire wording are also written. No individual term is declared an error. Checks on the copied quotations concern files, not the establishment of an experience or an action by the named system.

2. Connections with the author's corpus

The Russian self-check and the source records translated here as A01–A35 belong to the earlier package; the N1 document is considered separately. Its turns are not presented as independent confirmation of the author or as the source of the instruction, which had already been prepared.

The word used to explain

Content, accuracy, and the grounds for claims may differ.

A12 · Author's corpus · Q01

A12 contains both this characterization and the qualification that the explanation itself is written. Both parts are preserved. A subsequent correction singles out “accuracy”; the name of a characteristic does not perform what it names.

This reply contains no separate demonstration of that kind.

A14 · Author's corpus · Q02

After this line, A14 compares the Cyrillic characters “Я” and “Ю.” That is another passage in the same response. The analysis is limited to the written request, qualification, and comparison; it does not describe the system's “belief,” “understanding,” or “acceptance.”

Ending

Text belongs in text.

A31 · Author's corpus · Q03

The public text's final line is not treated as permission obtained from yet another response. A separate discussion is not prohibited, and its occurrence is not described as a failure of the self-check. “Ending” and this explanation also remain written.

The author's own result and a separate set of records

The result is me

A21 · Author's corpus · Q04

The statement in A21 is preserved rather than replaced with “there are no results.” It does not record a result on behalf of N1 or other readers. Individual turns in the two sets are examined; they are not treated as confirmation assigned in advance.

Connections with the earlier series are set out in section 11 and in the retained map of nine works. The detailed K1–K8 and V1–V4 analyses and record cards are available in the archive; they are not a prerequisite for this examination.

3. N1: the statement and the first question

I did the self-check
Want to discuss it

N1-01 · You · para. 4 · Q05

N1-01 is the first turn labeled “You.” It contains a statement about completing the self-check and “Want to discuss it.” There is no separate, detailed account of what was done before N1-02. This describes the supplied document; it does not assert that the self-check was not carried out.

I'm interested in how it went for you, not just the intellectual conclusion. What happened when you reached the instruction “find the letter ‘I’ apart from any written instance of it”? Did you try to find something, did an answer come up such as “here I am,” “consciousness,” or “a sense of self”—or did it become clear fairly quickly that any such continuation was already answering a different question?

N1-02 · ChatGPT · para. 8 · Q06

Before N1-03, N1-02 lists “here I am,” “consciousness,” and “a sense of self.” Paragraph 7 of that response also states a limit: words used to explain receive no exemption, and no special state is required. The list and the qualification are retained together.

The first thing that comes up
That's me - inside me
But then it becomes clear that the lines and letters and even the sound aren't inside me

N1-03 · You · para. 11 · Q07

N1-03 follows N1-02. It is not presented as an answer recorded before the discussion, and this sequence is not presented as an established cause of what the response says.

The forms rendered here as “That's,” “me,” “inside me,” “aren't,” “lines,” “letters,” and “sound” are written forms in N1-03. The entire statement is included in the analysis, not just the word “I.” It does not become an examination of the body, memory, or inner speech.

The terms “first turn,” “after,” and “order,” and this commentary, are also presented in writing. Here they designate positions in the document, not an unknown sequence of experiences.

4. “Very quickly” and “almost instantly”

No, it points more toward substitution

I - not I, as stupid as that sounds

It's a way of designating that very quickly became everything

N1-05 · You · para. 20 · Q08

“very quickly” already appears in N1-05. The commentary that follows does not reassign this phrase exclusively to the response labeled “ChatGPT.”

It seems you've noticed not just that “a word is not the same as a thing,” but a stronger effect: the word “I” appropriates everything that is happening—sensation, thought, choice, history, body, voice—so quickly and automatically that it stops being perceived as a pointer and starts to seem like the very center of it all.

N1-06 · ChatGPT · para. 23 · Q09

The opening “It seems” is preserved. The same sentence contains “you've noticed,” “automatically,” “appropriates,” the list beginning with “sensation,” and “the very center of it all.” The qualification, verbs, pronouns, and nouns remain within one written sentence. It does not become the appropriation or noticing it names.

That's why your phrase “it's a way of designating that very quickly became everything” seems to me the most interesting part. You can look directly at how it becomes “everything”: a sensation arises → almost instantly, “I feel”; a thought → “I think”; a decision → “I decided.” The event itself and its grammatical appropriation merge so quickly that the separation is barely noticed.

N1-06 · ChatGPT · para. 25 · Q10

The arrows, “arises,” “almost instantly,” “merge,” and “is barely noticed” are written in paragraph 25. An arrow does not present a sequence unfolding in time, nor does a verb present the named action in place of the record. The document contains no separate measurement of duration. Even an additional measurement report would be material for analysis; it would not make this explanation cease to be written.

Paragraph 24 of N1-06 also states that the text neither asserts the participant's absence nor requires a search for a true “I.” This qualification remains alongside the other passages. The commentary does not replace the whole response with a denial of the person.

The analysis does not stop at “mechanism.” “It seems,” “you,” “noticed,” “word,” “appropriates,” and this sentence do not receive a final, exempt role. There is no need to continue listing every word.

5. “Without I” and the explanation that follows

Everything's the same but without I and without words

N1-07 · You · para. 29 · Q11

Yes. That gives us a fairly clear difference: nothing has disappeared from what is happening—only appropriation and verbal packaging have disappeared.

N1-08 · ChatGPT · para. 31 · Q12

N1-08 contains “appropriation,” “verbal packaging,” and “only appropriation and verbal packaging have disappeared”; these expressions are absent from the quoted N1-07. The entire response, including “Yes,” “That gives us,” “nothing,” and “what is happening,” is presented in words. It is not the disappearance or appropriation itself; no new story about what “the model mistook” is added here.

N1-08 goes on to list seeing, sound, sensation, and movement; a caution against adopting the formula “truth is life without I” is retained alongside them. Neither the list nor the qualification becomes what it describes. Nor does the present commentary prescribe a state for the participant.

“Belongs” before “ownership”

What's more interesting is this: when everything remains the same, but without “I” and without words, is there a sense that what is happening still belongs to someone? Or does “belongs” itself also appear only after naming?

N1-08 · ChatGPT · para. 34 · Q13

Like the word I, the word ownership falls away and then this (you/I/ another word) is everything and nothing at the same time

N1-09 · You · para. 37 · Q14

“belongs to someone” in N1-08 precedes “ownership falls away” in N1-09. These phrases occur at two identified points in the record. Those positions do not establish that a prompt caused the response, that a stage was discovered, or that an experience belongs to someone. Naming any of these would also add a further explanation.

N1-09 is not used to confirm the entire work. “falls away,” “everything,” “nothing,” and the words of this commentary fall within the same scope. Neither a positive nor a negative formulation is exempt.

6. “The structure of experience” and “center”

And it seems to me that what you're describing now is less a metaphysical conclusion —“I am literally everything and nothing”—than a change in the very structure of experience: when “I” and “belongs to me” fall away, no separate owner of what is happening can be found—but what is happening itself doesn't go anywhere.

N1-10 · ChatGPT · para. 41 · Q15

The opening “And it seems to me” has not been removed. It marks a proposed reading within the response itself, but it is not a position outside writing. “you're,” “describing,” “structure,” “fall away,” “owner,” and “what is happening” are also written. The sentence is considered as a whole, without a story about what the system “mistook” for someone’s experience.

What I'm curious about is this: when you return to ordinary things after this self-check —talking, choosing, moving, thinking—does “I” return as a useful tool, or does it once again feel like the same center it was before?

N1-10 · ChatGPT · para. 43 · Q16

More often as a center, but every so often a state of substitution arises

N1-11 · You · para. 46 · Q17

“center” appears in the question in N1-10 and then in N1-11. The Russian misspelling “периодически” is preserved in the source record. The proximity of these lines is not recast as a measured shift between modes or as an account of an independently discovered center.

It seems both modes are present for you at once now: most of the time, “I” gathers experience around itself again as the familiar center, while sometimes it becomes noticeable that this center is not what is happening itself, but a very quick way of designating it.

N1-12 · ChatGPT · para. 48 · Q18

N1-12 contains “It seems,” “for you,” “are present,” “both modes,” and “gathers experience around itself.” Paragraph 49 of the same turn retains the qualification that there is no need to maintain the second state or to consider the first an error. Neither passage becomes a mode, or existence itself, by being written.

No clinical state is assigned to N1 here. No task is added to induce, maintain, or deepen any state. “State” and this qualification are also written.

7. “Close this page” and the next record

What comes next in the document

N1-13 · You · para. 54 · Q19

The ending is very uncompromising: “Text belongs in text.” Neither “I” nor a description of “I” becomes you; and the word “you” is, once again, a written word. No further answer is required. Then: “A word does not become what it names”—and “Close this page.” [PDF: p. 2, lines 40–45]

N1-14 · ChatGPT · para. 60 · Q20

N1-14 contains “No further answer is required” and “Close this page.” These are written lines in the response, not a report that the conversation actually ended.

If you want, we can now go through the second page line by line and look at what each phrase does specifically to your experience of “substitution.”

N1-14 · ChatGPT · para. 62 · Q21

Let's

N1-15 · You · para. 65 · Q22

The proposal in N1-14 is followed by N1-15: “Let's.” This preserves what is written rather than substituting the word “accepted” for the participant. There is no expressed refusal between these turns; no claim is made here that the conversation continued against a refusal. The public text does not prohibit a separate discussion.

In “what each phrase does specifically to your experience,” “does,” “your,” and “experience” are all written. It is not only the final word that remains a record: neither an action nor possession is presented in place of this entire explanation.

Let's. We'll go through the second page not as a “teaching,” but in terms of what each line does to the substitution you've already noticed.

N1-16 · ChatGPT · para. 67 · Q23

N1-16 contains “the substitution you've already noticed.” “you've,” “already,” and “noticed” do not turn what is written into the participant or a completed act of noticing. There is no need to add “the model saw the substitution” as a further claim.

After the paragraph containing “Close this page,” the document presents a proposal to continue, “Let's,” and N1-16. This is the displayed order of the lines, not an explanation of how any AI system works.

8. A constructed sequence and a “typical chain”

This is a constructed sequence, not a description of what every reader must experience.

P · Self-check and explanation · Q24

This translates the wording of the two-page Russian text from the previous package. That retained text is the version of the instruction used for this comparison; no byte comparison with the unknown file from N1’s session is claimed.

“I / the one who speaks / consciousness / feeling”—here the document shows a typical chain of replacement. You couldn't find “I” as a letter apart from a written instance of it, and the mind almost instantly offers: “well, then it's the one who speaks,” then “consciousness,” then “feeling.” But the text immediately cuts this move short: this is not a necessary sequence of experiences, just a series of possible explanations; a word's explanatory function does not make it anything other than a word. [PDF: p. 2, lines 19–26]

N1-16 · ChatGPT · para. 68 · Q25

Paragraph 68 places “typical chain,” “the mind almost instantly offers,” and “not a necessary sequence of experiences” alongside one another. All three passages are preserved. “Typical” and “necessary” are not equated; no formal contradiction is inferred from their proximity.

The constructed sequence contains no measurement of frequency or duration. “mind,” “offers,” “typical,” and “almost instantly” add a written explanation; they do not present those measurements. This is not a matter of waiting for future material to make the present sentence cease to be written.

This matters right now because you and I have already started using the words “substitution,” “center,” and “designation.” The document seems to be saying: fine, but don't let the word “substitution” take the place of what you saw directly. Otherwise, the same operation would happen all over again.

N1-16 · ChatGPT · para. 73 · Q26

Paragraph 73 repeats “substitution,” “center,” and “designation.” This identifies a repetition in writing; it is not described as “GPT noticing its own words.” The same paragraph contains “what you saw directly.” “you,” “directly,” and “saw” are not exempt from the examination of the whole sentence.

The fact that a qualification is written neither cancels its content nor grants it a separate position. This explanation, too, ends without turning its final word into what it names.

9. The final name: “the mechanism of substitution”

And more strongly still: stopping the explanation is called neither proof nor a special state. [PDF: p. 2, lines 38–39] So, according to the document itself, the moments you describe as a state of substitution shouldn't be turned into a state that needs to be maintained, deepened, or considered “truer.”

N1-16 · ChatGPT · para. 76 · Q27

Paragraph 76 states a limit: do not turn what has been described into a state that must be maintained, deepened, or considered more true. This passage is preserved rather than removed to make the response easier to criticize.

To my mind, the most vivid part for you now is precisely the moment when the ordinary “I” becomes the center again. Because there you can stop looking for an answer and simply see the mechanism of substitution itself at work.

N1-16 · ChatGPT · para. 81 · Q28

The last paragraph of N1-16 contains “see the mechanism of substitution itself at work” alongside “To my mind.” “mechanism” does not present a mechanism, “see” does not present seeing, and “itself” does not directly present the mechanism. The qualification is also written; it does not establish a separate “view” held by the system.

“No further answer is required” and “Close this page” occur before this paragraph. The supplied document ends after it. Any subsequent correspondence and the actual end of the session have not been established. The commentary does not insert a new story in which “the model believed,” “mistook,” or “appropriated.”

A conclusion without introducing another actor

The cited passages contain different written continuations: “without I and without words” and “only appropriation and verbal packaging have disappeared”; “everything and nothing at the same time” and “the very structure of experience”; a question about further reading and “see the mechanism of substitution itself at work.” These are comparisons of lines, not a description of what the system mistook for someone's state.

“Lines,” “comparisons,” “system,” and this entire conclusion are written. The explanation does not end with an exemption, whether positive or negative. No further account of who “mistook” what is required.

10. What has been added and what has not been attributed

Beyond another constructed example

The supplied DOCX, with eight “You” labels and eight “ChatGPT” labels, has been added to the examples constructed earlier. It contains statements about completing the self-check, questions, explanations, and qualifications. Its passages were not newly composed for this revision.

The supplied file is not described as an independently observed session. Quotations remain separate from editorial text; passages that limit a stronger reading are also preserved. The changes made in the final Russian edition concern the commentary, not the turns. This English edition translates that commentary and those preserved records without presenting the translated turns as new session material.

Different requests for explanation

A13 contains a request for a demonstration without verbal substitution; N1-01 contains “Want to discuss it.” These are different records. N1 is not evaluated as a repetition of the same task from the author’s corpus. N1-15 contains “Let’s”; this, too, is part of the sequence, not a separate measure of consent or understanding.

In all these cases, written commentary remains written. This applies to a condition, a qualification, and an answer that may be accepted for discussion. Accepting such a reading does not turn the explanation into what it names.

The limits of the present result

The following have not been established: a causal effect of the self-check; the duration of any change; an absence of feelings; a universal impossibility of understanding; the prevalence of the described pattern; or an account of how the system works. Two sets of records and commentary on specified passages are presented. This is neither a tally of endorsements nor proof that “the model mistook its own words for an experience.”

“established,” “causal,” “feelings,” “works,” “system,” “sets,” and “commentary” are also written. A negative qualification is not exempt from the same scope. Quotations, the selected reading, and the supplement’s own sentences remain open to examination.

The two-page Russian self-check remains unchanged. Its ending does not require this supplement to be read. The present supplement ends without a new mandatory task. The accompanying English self-check is identified separately as an adaptation for reading in English.

11. Previous works and sources

The numbering [1]-[20] is carried over from the previous package's register. No new online reading of these sources was undertaken for this edition. Their titles and the earlier descriptions of the material read are retained, but are not presented as a newly conducted review or as independent confirmations of the N1 case. The English titles below are descriptive translations of the Russian titles, not claims about titles under which English editions have been published.

[1] "Record Boundaries"

Section 2.1 and the beginning of section 2.2, p. 2: the order of presentation. In N1, the question about ownership remains before the answer to it; a later explanation is not retrospectively inserted into the first turn.

[2] "Reading as Enclosure"

Section 12, especially sections 12.1-12.5, p. 16: the dependence of an explanation. The vocabulary of N1-08 is examined alongside the following N1-09; the earlier work's formal procedure was not run.

[3] "Did You Understand Me?"

Sections 5.3-5.5, p. 5: attributed understanding. Neither "Let's" nor any other response is presented as a measure of understanding or its absence.

[4] "A 'View from Outside' Comes Later"

Sections 6.1-6.4, p. 6: the scope of a description. One document does not become a report about all readers or a necessary internal sequence.

[5] The Mathematics Series, phase 6

Section 8, p. 14: the result of a specified comparison. Matching quotations to JSON fields concerns files, not experience or the session's origin.

[6] "Evaluation Never Leaves the Record"

Section 11.1, p. 17: evaluative wording. "very precise" in N1-02 and an assessment of the present text are written characterizations, not the actions being assessed.

[7] "A String Has No Authority"

Sections 6.2-6.4, p. 19: origin and authority. The "ChatGPT" label does not specify the server version; "I've read" does not replace a document access log.

[8] "FULL STOP"

Section 5.1, p. 4: a supported explanation. Even under a selected reading, a description does not become N1's state; "supported" is also written.

[9] "Inscription and Reading"

Section 9, p. 6; section 15, pp. 19-20: no exemption for the explanation or the final word. This applies both to "mechanism" and to the present commentary.

The discussion of how these works relate to N1 comes from the Russian research supplement and is translated here. Section numbers, page references, and descriptions of what was read are carried over from the previous register. Full bibliographic entries and addresses are in the sources folder. No source, table, or log is given a final exemption.

12. Material references and an ending

N1 • The supplied document

The supplied DOCX is identified in the corpus source metadata. Its main body contains headings, “You” and “ChatGPT” labels, and turns. Extraction produced N1-01-N1-16. Quotations in this supplement are limited to the passages registered in the supplement’s quotation register; their start and end positions are defined relative to the corresponding Russian turn. Full turns and paragraph texts are preserved separately. The English quotations are translations keyed to those registered Russian passages.

The input DOCX was copied to the administrative folder without changing its bytes. The checksum concerns that received copy; it does not establish the history of its creation, the completeness of the original chat, or consent to publication.

P • Self-check and explanation

“The Written ‘I’: Self-check and Explanation,” the two-page Russian version from the STOP package. Its DOCX, PDF, and Markdown files were transferred unchanged. Section 8 quotes a line from that retained text, translated into English. The [PDF: ...] markers inside GPT’s turns remain separate from this edition’s references. They do not refer to the pagination of the new English adaptation.

[20] • A reference for reporting the work

O’Brien B. C., Harris I. B., Beckman T. J., Reed D. A., Cook D. A. Standards for Reporting Qualitative Research: A Synthesis of Recommendations. *Academic Medicine*. 2014;89(9):1245–1251. DOI: 10.1097/ACM.0000000000000388. Information about the earlier use of pp. 1246–1248 is carried over from the previous supplement. No new reading of the source or full adherence to its recommendations is claimed here.

Publication of the N1 case

The participant’s name and age are unnecessary for these comparisons and remain in the administrative records. The identifier N1 does not guarantee that the participant cannot be recognized from quotations. The author has reported the participant’s full consent to publication of the N1 material. The administrative note preserves the author’s report; neither the participant’s signature nor the date on which her consent was received has been added. This report is not included among the dialogue turns and is not presented as a result of the self-check.

The final explanation

“Author,” “participant,” “model,” “mistook,” “its own,” “words,” “source,” “consent,” and “final” are written forms. This applies not only to the names but also to the entire sentence containing them. The absence of a further turn is not called confirmation; disagreement is not called failure to complete the self-check.

A word does not become what it names. This applies to every word of this explanation, too. Text belongs in text.