# Writer-Defined AI Personas for On-Demand Feedback Generation

Karim Benharrak
karim@cs.utexas.edu
The University of Texas at Austin
Austin, TX, USA
University of Bayreuth
Bayreuth, Bavaria, Germany

Tim Zindulka
tim.zindulka@uni-bayreuth.de
University of Bayreuth
Bayreuth, Bavaria, Germany

Florian Lehmann
florian.lehmann@uni-bayreuth.de
University of Bayreuth
Bayreuth, Bavaria, Germany

Hendrik Heuer
hheuer@uni-bremen.de
University of Bremen
Bremen, Bremen, Germany

Daniel Buschek
daniel.buschek@uni-bayreuth.de
University of Bayreuth
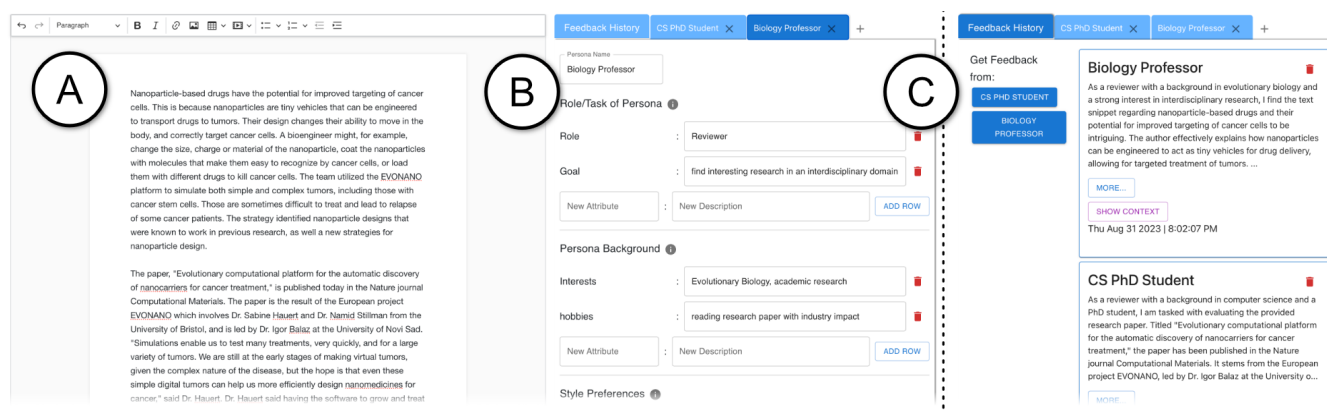Bayreuth, Bavaria, Germany

Figure 1: Our text editor *Impressona* supports writers with generated feedback on their draft: While working on a text (A), writers can define AI personas that represent their target readers (B). Writers can select any part of their text and click on a persona button (C) to receive feedback from the perspective of that persona. This feedback is generated with a Large Language Model (GPT-3.5) by prompting it with the provided persona information. (B) and (C) show different tab states of the same sidebar, not two sidebars.

## ABSTRACT

Compelling writing is tailored to its audience. This is challenging, as writers may struggle to empathize with readers, get feedback in time, or gain access to the target group. We propose a concept that generates on-demand feedback, based on writer-defined AI personas of any target audience. We explore this concept with a prototype (using GPT-3.5) in two user studies (N=5 and N=11): Writers appreciated the concept and strategically used personas for getting different perspectives. The feedback was seen as helpful and inspired revisions of text and personas, although it was often verbose and unspecific. We discuss the impact of on-demand feedback, the limited representativity of contemporary AI systems, and further ideas for defining AI personas. This work contributes to the vision
of supporting writers with AI by expanding the socio-technical perspective in AI tool design: To empower creators, we also need to keep in mind their relationship to an audience.

## CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **Text input**; • **Computing methodologies** → **Natural language processing**.

## KEYWORDS

Writing assistance, Personas, Text feedback, Large language models, Human-AI interaction

# 1 INTRODUCTION

People write *for* people [25]. Doing so well is challenging, as many readers of this paper will have experienced as writers themselves: Internal ideas need to be translated into a sequence of words to be brought to the page for the readers [17]. For instance, researchers of any field need to convey their findings to their academic community, including peers and reviewers, but also policy makers and the general public, in reports, magazines, blogs, and so on. Here, writing excels if it conveys information in such a way that it is easily understandable and relatable to the target group.

Despite this social nature of writing [57], writers are frequently left disconnected from their readers. Existing writing tools address this only to a limited extent: For example, the text checking tool *Grammarly* includes an "audience" setting (three levels of expertise) [1]. Already in 2010, Bernstein et al. [6] proposed a different approach with *Soylent*, a text editor that allowed writers to delegate tasks to crowdworkers. More recently, systems such as *Wordcraft* [64] or *CoAuthor* [41] provide co-writing experiences with AI. None of these approaches bring in the perspective of the target readers.

Doing so is challenging: It can be difficult for writers to get access to a target group during writing sessions, in time in between sessions, or at all. Thus, writers are rarely able to receive feedback "on-demand" from a reader perspective.

The goal of this paper is to address this and support writers in receiving feedback from the perspective of their target group through technology. As our design inspiration, we note parallels to user-centred design. Just as writers write for a specific group of readers, designers create a digital artefact for a target group. This target group may simply live in the back of the designer's mind but it is often beneficial to engage with it more concretely, for instance, via user research and the creation of personas.

We transfer this to writing and propose the concept of *AI personas for feedback*: Writers define personas of their target readers, which then provide feedback on the text. This feedback is generated by a Large Language Model (LLM) based on a prompt that includes the persona information. We investigate this concept with three research questions:

- **RQ1:** How do writers perceive the idea of AI personas?
- **RQ2:** How do writers define and use AI personas?
- **RQ3:** How well can contemporary LLMs support AI personas?

To address these questions, we developed *Impressona*, the first text editor that enables writers to create AI personas of their target readers to receive feedback (Figure 1). While working on a text (Figure 1A), writers first create a persona and enter its attributes (Figure 1B). Examples for such attributes include the persona's background knowledge (e.g. "computer science basics"), job (e.g. "professor") and role (e.g. "editor"). They can create multiple personas and thus reader perspectives. Writers select parts of their text and click on a persona to generate comments (Figure 1C). Our backend prompts OpenAI's GPT-3.5 to generate feedback, with the selected text and the persona information embedded into a prompt template. For an example of one such interaction moment from our user study, see Appendix A.2.

We use our prototype in two user studies (N=5 and N=11), with a design iteration in between. Participants brought their own writing projects to our study to work on real-world texts that are personally relevant to them. They appreciated the on-demand concept and strategically used personas for getting different perspectives. However, they found it challenging to initially define personas. While the feedback was often verbose and unspecific, it was seen as helpful and inspired revisions of text and personas.

In summary, we contribute: (1) A novel concept for writers to receive feedback, generated based on writer-defined AI personas that represent target readers, (2) its implementation in a functional prototype, and (3) insights from an evaluation with users.

In a broader view, we contribute to the vision of supporting writers with AI by expanding the socio-technical perspective in AI tool design: To support creators, we also need to keep in mind their relationship to an audience.

# 2 BACKGROUND & RELATED WORK

We clarify similarities and differences between personas in user-centred design and our concept of AI personas. We further relate our work to research on writing and writing support.

## 2.1 From Personas in UCD to AI Personas of Readers

A persona in user-centred design (UCD) [13] is a descriptive model of a group of users that captures information about how they behave and think, and what their motivations, needs and goals are. They help designers to avoid "self-referential design". Analogously, our goal is to help writers empathise with their readers and avoid "writer-based prose", which stays in the writer's own information organisation that might not be useful for readers [17].

The key difference to UCD is that writer-defined personas are likely not informed by user research. We also do not expect writers to be familiar with the concept of personas in UCD. Their "personas" of readers may nevertheless be grounded in experience – e.g. from past feedback by readers, by genre conventions, and so on. Recent HCI research suggests that writers indeed develop a mental model of whom (or which system) to approach for what kind of feedback [19]. Similarly, writing research discussed the concept of a "mental convention model" [20] that writers internalise as they become familiar with a writing context (e.g. genre, target community). These findings lend credence to the idea that writers can describe aspects of their target readers to generate feedback on.

## 2.2 Using Large Language Models for the Representation of People

Related work leveraged Natural Language Processing (NLP) to represent or "simulate" people [46]. A key line of work refers to "personas" as a foundation for generating text with language models from a certain perspective (e.g. [52, 53, 66, 67]), with HCI applications particularly for chatbots (e.g. [3, 21, 51, 59]). These personas are often short text descriptions. A related concept is the "system instruction" prompt, which tells an (instruction-tuned) model

which perspective to take (e.g. "You are an assistant that speaks like Shakespeare"[1]).

Related to writing, Nilsson et al. [45] introduced the concept of semantic avatars of (historic) authors based on a language model. More recently, Hämäläinen et al. [22] used an LLM to generate synthetic responses for HCI experiments. Similarly, Park et al. [47] used LLMs to generate user content during prototyping, such as comments for a discussion board. They included short text statements as "user personas" in their prompts (e.g. name + job).

While systems based on LLMs have received attention for their ability to generate human-like language [49], researchers have also pointed out their limitations. For instance, Kabir et al. [33] find that more than half of ChatGPT's responses to software engineering queries contain inaccuracies (52%).

These potentials and limitations motivated us to explore how LLM-based AI systems can support writers in practice, specifically through the lens of reader personas.

## 2.3 Writing Research & Practice: Writing for an Audience

Thinking about the target readers is common advice in writing guides. Rechenberg [48] motivates this as a safeguard against overly abstract writing. Moreover, he highlights that writers should continuously "talk" to an imagined inner reader to help them think about their readers' existing knowledge and known terminology, and anticipate questions. Similarly, Langer et al. [39] describe the concept of "person-centred writing". One of its key aspects is *empathy* – striving to imagine the readers' world of experience. We aim to support writers in doing so, by helping them to externalise the inner reader as a persona and by using AI to let it "talk".

Related, Göpferich [20] highlights the "communicative function" of texts in her framework on evaluating comprehensibility. One of the framework's dimensions is the "target group", with features such as age, gender, cultural background, education, hobbies, and prior knowledge, as well as the heterogeneity of the group(s). In our prototype, writers can describe reader personas along such dimensions, and create multiple different personas.

If writers should think about an inner reader, when do they think about the text? The different cognitive processes in writing are non-linear, as described by Flower and Hayes [17, 24], involving planning, translating, and reviewing. Writers consider the target audience on several levels of planning, from concrete content goals and plans (e.g. what to say/do to an audience), to middle-range goals (e.g. "appealing to an audience"), to top-level goals (e.g. "interest the reader"). Writers work on ideas and plans towards these goals through a pattern of exploration and consolidation. Our fundamental UI design accounts for this: Writers can easily switch from writing in a *page view* to a *sidebar* for capturing their understanding of the audience by creating and editing personas.

## 2.4 Designing Writing Tools: Expanding the Perspective Beyond the Writer

Writing tools have a long history [37, 42]. NLP capabilities are increasingly integrated into them, both in industry (e.g. "Copilot"

---

[1]https://help.openai.com/en/articles/7042661-chatgpt-api-transition-guide

in MS Office [54]), as well as in the HCI research community and its prototypes (e.g. see [12, 18, 29, 40]).

A key observation here is that most writing tools share a common perspective on the socio-technical environment that is relevant to their design: There is one person writing, to be supported by technology. As recently shown, acceptance of such support is impacted by social dynamics [19]. Related, AI-mediated communication [23] examines sender and receiver experiences of text written with AI (e.g. [28, 31, 44]). Together, these observations motivate us to explore a design perspective that considers that most writing serves to communicate information to others.

A specific related example is provided by recent research into writing support for experts for plain and simple language [26, 27]: These specialist writers found it important to have their text evaluated by stakeholders and felt that technology could support such reviews. We provide a first step towards supporting such feedback loops with AI.

## 2.5 The Role of AI Tools and Feedback from Other People when Writing

Writing support can come not only from AI tools, but also from peers. Gero et al. [19] interviewed creative writers about both options. They report that writers "learn specific characteristics of people that would modulate when they'd turn to these people for support." They also discuss individuality: AI reflects assumptions and a perspective based on – and biased by – its training data, which is different from an individual human. In this light, they call for "[m]aking explicit these assumptions, as well as highlighting when a model is trained to produce a different perspective". Related, recent work showed that a biased language model can influence viewpoints expressed by writers [30]. This is particularly noteworthy considering concerns about which human values are represented by LLMs [32]. In this light, we explore how writers might be *supported* by perspectives provided by a (pretrained) LLM, when shaped through prompts based on the writer-defined persona information.

## 3 CONCEPT AND IMPLEMENTATION

We introduce the concept of AI personas of readers, a novel approach inspired by combining personas from user-centered design with insights from research into writing and AI writing support systems (see Section 2). To explore this concept, we developed a prototype text editor called *Impressona* (Figure 1).

### 3.1 Concept Development

We developed the concept and prototype based on insights from the literature and a design iteration with feedback from users.

*3.1.1 Design Iteration.* We developed the prototype in two iterations. First, we implemented an initial prototype (V1) and tested it with five participants. Findings from this first study round motivated these three UI changes for V2:

- *Persona definition:* V1 had a single field for free text input. Our findings motivated more guidance here. Thus, we changed

this into a form-like view in V2, to give it structure (attribute-description pairs). We also added an info button that shows examples.

- *Multiple personas:* One persona at a time was not enough. We therefore changed the prototype in V2 to allow for creating multiple personas and switching between them via a tabs view.
- *Feedback history:* Participants in the first study wished to re-view previously generated feedback, but V1 only showed the last feedback. This motivated us to integrate the "Feedback History" in V2.

In the rest of the paper, we always refer to the second version (V2), if not stated otherwise as V1.

*3.1.2 Design Rationale and Goals.* We extracted the following higher-level design rationale and goals from the literature and insights from our first user test with V1: 1) As the key idea, *supporting on-demand feedback* allows writers to receive feedback from a reader perspective while writing, through AI personas. 2) *Enabling writers to define personas themselves* allows them to express their target readers' characteristics, and to obtain different perspectives in this way. 3) Related, *supporting easy iteration of personas* enables writers to refine them and also to react to potentially unwanted aspects in the generated feedback. 4) *Supporting the definition and comparison of multiple personas* facilitates gaining insights through contrast, both into the draft, as well as into the AI's capability to manifest diversity in its generations. 5) Finally, *supporting local feedback* allows writers to request feedback on any piece of text (which we realized through text selection) to facilitate specificity of feedback.

## 3.2 Frontend

The frontend of our system is implemented as a web application using React to achieve the mentioned design goals.

*3.2.1 Main layout: Editor and sidebar.* The UI is split into two main areas – a text editor (Figure 1A) and a sidebar (Figure 1B). The sidebar's navigation panel at the top consists of a "Feedback History" tab for requesting feedback, one tab for each created persona (to edit it), and a "+" button for adding a new persona. This side-by-side layout supports the mentioned design goals via easy access to both the draft and the personas.

*3.2.2 Persona tabs: Create and edit personas.* This tab view supports the two design goals of enabling writers to define personas themselves and to easily iterate on personas. Each persona tab leads to a view with form fields (Figure 1B). Writers can give their persona a name that also shows up in relevant other places (tab header, feedback by that persona). The rest of the form is structured in rows of pairs of text fields (for attributes and descriptions, e.g. "job: physician"). Rows can be added and removed by clicking on respective buttons.

This form has four sections: "Role/Task of Persona", "Persona Background", "Style Preferences", and "Content Preferences". An info button in each section header provides guidance (Figure 2). Clicking on these info buttons presents users with a short description and a set of examples of attribute-description pairs that could



**Figure 2: A part of the UI for defining a persona: Hovering over the info button next to each category (e.g. "Style Preferences") provides a corresponding description, alongside examples of attribute-description pairs to facilitate the persona creation process.**

be added to this section. Writers can choose what to use – it is not required to add information to all categories.

*3.2.3 Feedback tab: View and request feedback from the personas.* The "Feedback History" tab (Figure 1C) realizes the main design goal of providing on-demand feedback, as well as the goal of allowing for comparisons of feedback from multiple personas. It holds a list of interactive cards that display the generated feedback. New feedback is displayed at the top. Cards start in a collapsed state that cuts off the text after a few sentences. A "See more" button allows users to expand the card to view the full feedback. A "Show context" button at the bottom of the card can be clicked to show the part of the text that the feedback was based on.

*3.2.4 Interaction.* We realized three interactions with the personas:

*Persona creation.* Writers click a "+" button in the sidebar's navigation panel to create a new persona. They then define their persona by filling out the form fields in its tab.

*On-demand feedback.* To request feedback from a created persona, writers first select the part of their text they want to receive feedback on (in line with the design goal of supporting local feedback). They then click on a persona's button in the "Feedback History" tab. This adds a new comment card in the feedback history, displaying the feedback.

*Persona refinement.* If writers are not satisfied by a response, or identified extensions or new directions for the persona, they can switch to the persona's tab at any time, to add, delete or edit fields; this supports easy iteration.
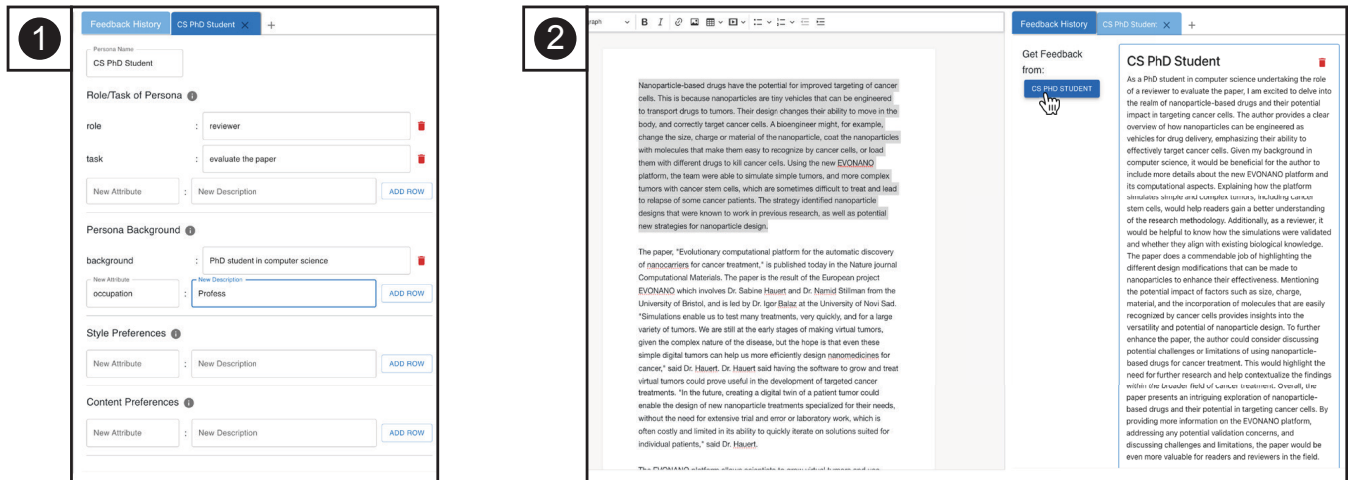
**Figure 3: The overall workflow: (1) The writer defines a persona by adding rows of attributes and descriptions in a persona tab. (2) In the text editor, the writer selects a part of the text and then clicks on a persona button in the sidebar to generate feedback from that persona for that piece of text.**

## 3.3 Backend

Our backend serves as a layer between our frontend and the OpenAI API[2] that we use to generate the feedback.

*3.3.1 Implementation.* We implemented our backend as a Python Flask API. Our backend receives the selected text and the clicked persona from the frontend. We use this information to create a prompt that we send to the OpenAI API to generate the feedback. Upon receiving the response, the backend sends it to the frontend, to be displayed in a feedback card.

*3.3.2 Feedback Generation.* We leverage OpenAI's Chat Completion using GPT-3.5 to generate the persona feedback. See Appendix A.1 for our prompt template. We provided six input-output examples with this prompt template, as few-shot examples for the model. This was based on our observations in early experiments to improve the model's adherence to the desired output. The feedback in these examples connected persona attributes to specific suggestions, to reveal how a persona's attributes affect the generated feedback (e.g. "As a history professor, I value that this text..."). These six examples were created from our own interactions with the OpenAI model, based on interactive sessions in which we explored the model's suitability at the start of the project (i.e. the few-shot examples are "cherry-picked" and manually refined input-outputs).

## 4 METHOD

We conducted two study rounds with our prototype, with a design iteration in between (Section 3.1.1). Both studies used the same methods – a writing task with think-aloud, a semi-structured interview, and a final questionnaire.

## 4.1 Apparatus

We used the prototype as described in Section 3, hosted as a web app: The first study used V1, the second study used V2. Besides

the features described earlier, V2 also implemented functionality to record user interactions with the system (e.g. creating personas, changing them, requesting feedback).

## 4.2 Participants

In total, 16 people participated in our evaluation (Table 1): five in the first study (2 male, 3 female; referenced as $P_1$ - $P_5$), using prototype V1; and eleven in the second one (7 male, 4 female; $P_6$ - $P_{16}$), using V2. Their age ranged from 22-33 years. All were proficient English speakers; four were native speakers.

We recruited participants from our networks across a few universities. Our goal was to find people with real writing projects. In the first study, we were open to include various kinds of texts, to explore the concept. In the second study, we focused on the academic context, as one relevant target group for new writing tools (cf. [56]). We report the resulting text types in Section 5.2. Participants' backgrounds included Master students and PhD students as well as researchers. Their median self-reported writing proficiency was 4 on a scale from (1) low proficiency to (5) high proficiency. The audiences participants regularly write for include *scientific community, myself, professors, general public, managers, and colleagues.* Participants in our second study were generally interested in topics around Artificial Intelligence (see Figure 8). Participants were compensated with € 15.

## 4.3 Procedure

The study was conducted remotely using video call software and screen sharing. The sessions were scheduled for 60 minutes and structured as described next (Figure 4):

*4.3.1 Study Intro (10 minutes).* An intro page on our web app explained the study in line with our institutional regulations (incl. GDPR, privacy and data collection info and informed consent). A researcher gave a live demonstration of creating and using personas

---

| P | Writing Project | Regular Audiences | Writing* | English** | Age | Gender | Occupation |
|---|---|---|---|---|---|---|---|
| P1 | Bachelor thesis, Email to Professor, Fiction Story | Creative Writing, Video Game Writing, Scientific Writing | 4 | 5 | 25 | Female | Student |
| P2 | Blog post (AI in education) | P2 themselves, professors | 4 | 5 | 26 | Female | Student |
| P3 | Research Paper (HCI) | Scientific community | 4 | 5 | 32 | Male | Researcher |
| P4 | Research Grant Proposal (Computer Science) | Researchers, general public, students | 4 | 4 | 33 | Male | Associate Professor |
| P5 | Blog post (Cooking) | P5 themselves, professors | 3 | 3 | 32 | Female | Student |
| P6 | Web article (Computer Architecture) | Academic audience | 5 | 5 | 23 | Male | Student |
| P7 | Business Report (ERP system in HR) | Managers, Co-workers, professors | 2 | 4 | 22 | Male | IT-Consultant |
| P8 | Research Paper (HCI) | Scientific reviewers, Management Teams, C-Level of coporates/start-ups | 4 | 4 | 29 | Female | PhD equivalent |
| P9 | Research Paper (HCI) | Human-Computer Interaction and Human-Robot Interaction audiences | 4 | 5 | 33 | Male | PhD equivalent |
| P10 | Research Paper (HCI) | Domain experts, researchers, colleagues | 5 | 5 | 32 | Male | PhD equivalent |
| P11 | Research Paper (Process Analytics) | Computer Science scientists | 3 | 5 | 27 | Male | PhD equivalent |
| P12 | Research Paper (HCI) | Scientific conferences, academia, work-related audiences | 5 | 5 | 25 | Male | PhD equivalent |
| P13 | Research Paper (NLP) | Academic audiences in the fields of NLP, Machine Learning, etc. | 4 | 5 | 23 | Male | PhD equivalent |
| P14 | Masters Thesis Abstract (AI in medicine) | None | 2 | 3 | 27 | Female | Patient management in hospital |
| P15 | Research Paper Summary (Canine psychology) | Uni (exams, course work, presentations) | 3 | 5 | 31 | Female | Student |
| P16 | Research Essay (English language variety) | Professors at university | 3 | 5 | 26 | Female | Student |

Table 1: Overview of the participants. *Writing proficiency, **English proficiency (both on Likert scales, 1/low to 5/high).

in our prototype, and encouraged participants to ask clarifying questions.

*4.3.2 Writing Task (35 minutes).* We asked participants to share their screens with recording. We explained the task of working on their text. During the recruitment, we had asked participants to think of writing tasks that they could bring to the study. Eleven people brought an actual writing project, 5 had come up with one for the study. A researcher observed the task and took notes. They were muted for most of the time, occasionally reminding participants to share their thoughts or asking a question to better understand people's thinking.

*4.3.3 Interview (10 minutes).* We conducted semi-structured interviews after the writing task to further learn about people's experiences. We asked these main questions: "What was your thinking when creating the personas?", "What did you think about the generated feedback?", "How did you use the feedback?", "What did you like about this idea?", "What could be improved/extended?"

*4.3.4 Questionnaire (5 minutes).* Participants provided their demographics and in study 2 completed a final questionnaire asking about their experience with the overall concept as well as specific aspects and UI features. It included Likert items on *Experience using the system* and *AI interest.* We took the latter from the "MeMo:KI – Opinion Monitor AI" [2]. These Likert items were constructed as statements reflecting the participant's viewpoint, such as "The feedback influenced my writing", "The feedback made it easier for
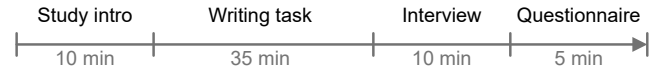


Figure 4: The study procedure, as described in Section 4.3.

me to write my text", and "I read articles about Artificial Intelligence with great curiosity". See Figure 8 for all items.

## 4.4 Coding of Think-Aloud, Interviews, and the Generated Feedback

The transcribed interviews and observation notes were analyzed using a German qualitative content analysis method [43] that is equivalent to thematic analysis [8, 9]. Using axial coding principles [14], the first author and second author individually reviewed the texts several times, moving back and forth through the material. They identified overarching categories/themes and subcategories of codes through an iterative process of clustering, splitting, and merging codes. The final themes were then discussed with the other authors in videoconferencing sessions. We repeated this process until a consensus was reached. In the same way, we also coded the generated feedback texts.

## 5 RESULTS

### 5.1 Overview

*5.1.1 Interaction (Logging Data).* We logged interactions in study 2: On average, participants created 2.36 personas (SD 0.48, range 2

to 3). They requested a mean of 6.27 feedbacks (SD 2.83, range 3 to 11), and revisited created personas 7.46 times on average (SD 5.70). Their final texts in the editor had a mean length of 438 words (SD 284.14). Figure 5 shows an overview of the attribute-description pairs that participants entered to define their personas.

*5.1.2 Perception of Interaction (Likert Data).* Participants in study 2 rated several Likert items at the end (Figure 8). In summary, these ratings overall indicate that participants found the feedback to be helpful and useful. Most rated the feedback as having impact on their writing, influencing content and/or wording. Despite this influence, almost all felt that they were the authors of their text. They also did not feel as if they would be interacting with real people. Participants were critical of the length of the feedback. Combined with the qualitative insights reported below, we learned that the feedback was often too long.

*5.1.3 Overview of the Generated Feedback.* By manually coding and analyzing the 68 generated feedback texts (Section 4.4), we identified a recurring structure with the following building blocks: In the **opening** sentences, the persona described its attributes, typically like this: "As a [e.g., reviewer, professor, journalist, ...], with [certain attributes], my task is [...]." This was almost always followed by a positive comment or summary of the user's selected text snippet, and often concluded with a suggestion on how to improve the text.

In the feedback's **main part**, the AI persona suggested adding more examples (47 occurrences), asked for topic-related content (26), sought clarifications (34) and more details (13), or gave specific hints, such as missing data, highlights, need for citations, or conclusions. In addition, the persona proposed improvements to the writing style (9), including suggestions for shorter sentences, easier language for accessibility, or different terminology. This advice was typically first described on a higher level, followed by a concrete suggestion (27), most often with a variation of "For example, instead of [text snippet], the author could write [proposed text]."

Finally, almost all feedback texts (66) concluded with a **summary**, often starting with "Overall, the text snippet..." or less frequently with phrases like "In summary..." or "In conclusion...".

Generally, we observed the AI personas to provide polite and uplifting feedback. Many times (63 occurrences), the feedback emphasized positive aspects of the text. Even when one participant ($P_{12}$) created a "mean" persona, although less encouraging, the feedback remained polite and included self-critical statements like "[...] the mean attitude in this context may affect the credibility of the feedback."

While the feedback's structure consistently followed the described scheme, its content was tailored to each participant's text snippet and the persona's preferences and background, rather than offering generic suggestions (e.g. see Figure 7).

## 5.2 How do writers perceive the idea of AI personas? (RQ1)

Participants used AI personas in various text types, including blog articles (3), business reports (1), research papers (6), grant proposals (1), research paper study notes (1), essays (2), and abstracts (2). Across all of them, they valued AI personas for gaining feedback

from different perspectives. For example, $P_9$ found it "nice to create multidisciplinary personas to have different perspectives". $P_{12}$ valued "to have, like, somebody, like, just write to you, OK, this needs more reasoning". And $P_2$ said that "it is important to know how a teacher who works at a school would go about this. What would her thoughts be?"

Eleven participants emphasized that it was a challenge to consider multiple points of view without AI personas, and saw the personas as a tool that helped them with that. For instance, $P_{13}$ said that "it's often hard to think of these things on your own" and "to look at it from different angles". Similarly, $P_4$ compared the tool to "a copilot to tell you, yeah, you forgot about this or that. That's good." And $P_6$ described that the persona feedback "[..] kind of makes you step outside of your writing", including "stuff I didn't notice before".

Five people drew explicit comparisons to other AI tools. For example, $P_{12}$ contrasted the concept with AI tools that give text suggestions: "It's much more valuable than just rewriting the sentence, because for that I could just use, ChatGPT, for example.". Similarly, $P_{15}$ said: "[the persona interaction] feels more like a person, like you kind of have this picture in your head of who you're talking to. It's a bit different with like generic AI models like ChatGPT [...]." When asked about what he liked about the concept, $P_{13}$ highlighted that "this layout and like being able to add personas [...] it's like a nice layer over whatever language model you're using underneath". These comments point at the perceived usefulness of the persona concept beyond LLM-based support in general.

## 5.3 How do writers define and interact with AI personas? (RQ2)

We observed certain strategies that writers developed when interacting with the AI personas.

*5.3.1 Types and Roles of Personas.* Writers defined different types of personas and used them for different roles and goals. Appendix A.4 shows an overview. A first type was expert personas. Participants defined experts – such as History Professor, PhD Advisor, Doctor in Radiology – to receive feedback, often on factual correctness. For example, $P_6$ explained: "I first defined a hardware expert to know what I was writing was correct and accurate".

Another type of personas was based on real individuals. For example, $P_{14}$ said: "I was actually imagining my professor that taught that course."

A third type was based on roles in the publication process, such as "Reviewer" or "Editor". For instance, $P_1$ said: "Does this even fit with the publisher I'm applying to? You know, if I'm writing youth literature for children aged 10 to 14, then I can adapt that accordingly and look here."

Another pattern was that four participants created personas for "strict" and critical perspectives on their text. For example, $P_{12}$ said "you wanna get the feedback from the mean reviewer".

*5.3.2 Interacting with Personas.* Three patterns emerged for interaction with the personas (as opposed to the text): Initial exploration, defining personas, and refining personas.

*Exploring AI Personas.* All participants found it interesting to explore the personas' capabilities. For instance, $P_8$ said: "I was

(a) *Attribute* words used in personas.



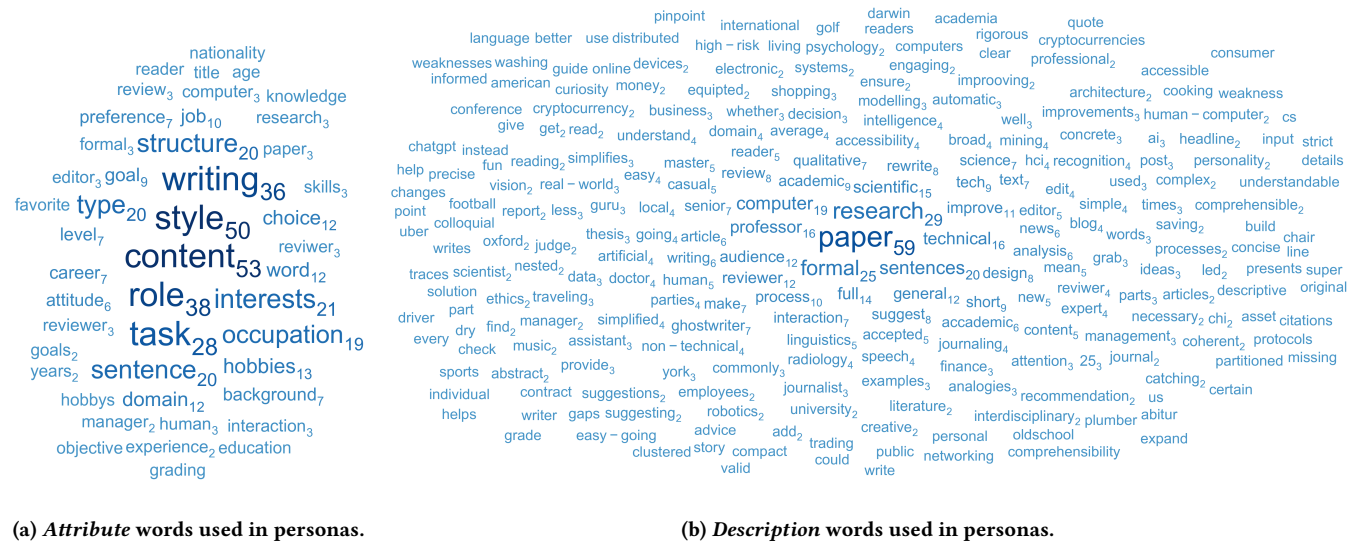(b) *Description* words used in personas.

**Figure 5: Overview of the words that participants entered in the attribute-description pairs (see UI in Figure 2), lightly scaled by how often these contributed to generated feedback (also annotated as numbers if >1). For example, if a participant defined "writing style: formal, scientific" for a persona and requested feedback four times from this persona, this would contribute counts of four to "writing" and "style" (in 5a), as well as to "formal" and "scientific" (in 5b). We removed stopwords (e.g. "the"). While shown as single words here for a better overview, participants also entered descriptions as phrases, not only single words (e.g. "Provide real-world analogies to ideas in the paper"). Log files including verbatim descriptions are available in the project repository (see link in Section 7).**

super curious to see, like, how the different personas would give a different kind of feedback". Sometimes participants were also just interested in exploring the concept through different kinds of feedback, such as $P_7$: "Might be funny to define a persona like "mum"".

*Defining AI Personas.* Initially defining personas was challenging for participants, even after our design iteration added more guidance in the UI in V2 (Section 3.1.1). For example, it was described as "not that easy" ($P_1$) and "actually hard" ($P_{12}$), and $P_{16}$ had "No idea, I'm not creative enough". This was also evident in the time participants took to create their initial persona's profile. On average, participants (using V2) required 3.82 minutes (SD: 2.17), with the swiftest definition completed in just under one minute, while the longest duration exceeded seven minutes for a participant to define all attributes.

A strategy that participants used to overcome this challenge was to define AI personas based on real individuals, which makes this task more concrete. For example, this enabled $P_9$ to use "[that person's] backgrounds, their research interests, probably even the keywords that they have mostly on the papers". Another pattern that helped them develop personas was iterative refinement, as described next.

*Refining AI Personas.* Participants iteratively refined personas: They created a persona, requested feedback, and read that to decide whether it is the kind of feedback they wanted. If not, they added, removed or changed attributes. However, it was not always obvious how to achieve this. $P_{10}$ wondered: "What should I change? Is that

level of granularity relevant?". This workflow also motivated three participants to suggest conversations as a way to refine the persona, such as $P_{12}$: "Can I iterate the feedback? I'm missing a chat here to refine. [...] It would be really cool to ask personas questions like *would you use such a tool?*"

### 5.3.3 Strategies for Writing with AI Personas.
We observed different strategic workflows that participants employed when writing with the help of AI personas. While we describe them separately, they are not mutually exclusive.

*Use of the Working Time.* Figure 6 shows how participants split their time between working in the editor and the sidebar (based on logged UI events): Some switched only a few times, focusing on either writing or engaging with the personas for longer periods. In contrast, others switched frequently. Across both these groups, most participants started in the text editor (e.g. copy-pasting in the writing project they brought to our study) and then switched to creating personas during the first third of the study. After engaging with their personas for the first time, participants asked for feedback approximately every 3 minutes and 37 seconds. The most engaged participant sought feedback every two minutes, while the least engaged waited 4 minutes and 45 seconds. For participants who switched more frequently, two patterns emerge: Some (4) seek feedback mainly from various personas on the same text snippet, while others (2) consistently use a single persona across different text parts. We found no relationship between frequency of use and the perceived helpfulness of personas.
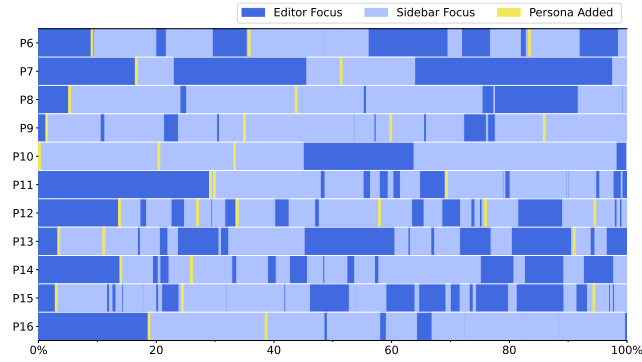
**Figure 6: Overview of participants' workflows (in study 2): Each row (y-axis) is one participant. The x-axis shows time as study session progress. Color shows participants' focus on the editor (dark blued) or sidebar (light blue), as derived from logged UI events. Yellow lines indicate when a new persona was added. Overall, some participants switched between writing and engaging with personas more frequently than others. See Section 5.3.3 for details.**

*Writing with AI Personas "In-the-Loop".* A key strategy that we observed among participants was an iterative "feedback loop". One example was $P_{11}$: After an initial feedback, they incorporated suggested changes into their text and requested another feedback from the same persona to verify whether they had improved the paragraph. We saw that writers were checking back with personas to see whether these were "satisfied" with the changes. For example, $P_8$ said: "Now I rewrite it myself and then I let the reviewer check it again and see if it actually made a difference". And $P_{10}$ said: "It seems like the change is reflected in the new review".

*Writing with Multiple AI Personas.* All participants from both studies used multiple personas. The key motivation was to gather different views. For example, $P_3$ "wanted it to be diverse" and $P_{13}$ wanted to "simulate different reviewers". A related strategy was to try and satisfy all these personas, by reflecting on and incorporating the feedback from all of them. Here, $P_6$ assigned different personas to different sections – a writing editor for their introduction with the task to reach a broader audience, and a hardware expert to check the main text. The Likert ratings also indicated that multiple personas were useful (Figure 8).

*Instructing AI Personas.* Participants were familiar with systems like ChatGPT. We observed that four participants used prompting strategies from such systems to further steer the feedback: For example, $P_{15}$ adjusted the "task" attribute of their persona with the prompt "find the original quote by [study author name]". And $P_9$ added an instruction prompt directly into the text for which they requested feedback (*"Below is the discussion of a research paper being written and I need feedback based on the structure"*).

*5.3.4 Impact of the Generated Feedback.* Participants found that the feedback helped them and had an impact on their text. We also observed impacts on the writers themselves.

*Helpful Feedback.* All participants (16 people) explicitly commented on the feedback as being helpful. The positive perception is also reflected in the Likert ratings (Section 5.1.2). For instance, $P_6$ said "But once it actually got to the advice, I really, really liked it, like it was super helpful.". The feedback was also deemed "valid" ($P_{10}$, $P_{12}$) and "correct" ($P_{11}$). It was also perceived as actionable: $P_{11}$ used the feedback "first, to actually review my writing process and second, to come up with new ideas" and $P_{14}$ mentioned that "I would have incorporated more [feedback] if I [had the time]. They had or suggested to use further research and more studies and things like that, and I definitely would have done that." An indicator of believable feedback is that it reminded $P_{10}$ "of reviews [from real people] I have read in the past" and $P_7$ believed the real person "would have answered pretty equal" to their created persona.

To further improve the feedback, participants expressed their wishes for additional features, such as knowledge of related literature ($P_8$, $P_9$) or longer context (e.g., by allowing for multiple selections of text ($P_{13}$) or considering the overall document context ($P_9$, $P_{16}$)).

*Impact of the Feedback on the Text.* Most participants (12 people) made changes to their texts in our editor based on the feedback they received. For instance, $P_6$ added a personal anecdote to their text after reading that the feedback suggested this. At the end, they said "I ended up making a lot of adjustments based off the feedback." Related, 9 out of 11 participants in Study 2 reported in the Likert questions that the feedback influenced their writing (see Section 5.1.2).

Moreover, four participants either kept generated feedback or edits resulting from using our tool for further use beyond the study. For example, $P_{12}$ copied one feedback into his own Overleaf project. And $P_4$ worked on a grant proposal and found that "this is actually quite useful, I will paste this paragraph into my proposal document. I can use it there."

*Impact of the Feedback on the Writer.* A different kind of impact was that for $P_6$ feedback resulted in pressure of incorporating suggestions: "I felt pressured to actually make the changes. Just because they're supposed to be like the expert in the scenario but it also felt strange to put their edits in exactly."

Observations and Likert results indicate that writers were influenced in their writing. At the same time, the Likert results show that most participants (9 of 11 people in V2) also reported that they feel to be the authors of their text (Section 5.1.2). As $P_9$ put it: "I feel like I have been the author of my text. I do not that much for ChatGPT".

Another finding was that we observed participants anthropomorphizing the system through their personas. For instance, $P_{16}$ made a sad face after being told that the writing time was almost up, saying "But I didn't talk to 'best friend' yet" before requesting feedback from that persona. Participants referenced their personas as if they could be real individuals. For example, they said "Let's see what X thinks about this", where X is the name of their persona ($P_6$, $P_{12}$). However, some participants (5 people) gave their personas generic names like "Persona 1".

## 5.4 How well can contemporary LLMs support AI personas? (RQ3)

Overall, we found that it is possible to use a contemporary LLM (GPT-3.5) to support writers with AI personas, albeit with some limitations and room for improvement, related to model and prompt engineering.

*5.4.1 Expectations about Contemporary LLMs.* As our participants were familiar with systems such as ChatGPT, some (3 people) stated that they recognize that the feedback was similar to that. For example, $P_{13}$ said: "Of course it looks a lot like something like ChatGPT". Related, some comments indicated that participants were interested in how their feedback is being generated. For example, $P_{11}$ noted that "It would be interesting to know what the data is it has been trained with". The prototype was also explored by contrasting different personas, which points into the direction of inquiring into the system through interaction (cf. "what if", scrutability [11, 34]).

*5.4.2 Quality of the Feedback.* As reported above (Section 5.3.4), participants found the feedback helpful. The Likert ratings (Section 5.1.2) support this as well, with a majority perceiving it as helpful, influential, and making it easier to write their text. Together, these results show that the model we used (GPT-3.5), and the way we prompted it, was suitable to explore the concept of AI personas.

That said, participants raised three key issues about the quality of the feedback generation. Concretely, they criticized it as verbose, unspecific, and repetitive.

The feedback was longer than needed. For example, $P_8$ said "it took me a while to read like the actual feedback". $P_{10}$ and $P_{15}$ suggested a "bullet point" format in contrast to the fully spelled out comments that were generated. Related to this verbosity, the feedback was partly perceived as unspecific, such as "very generic" ($P_9$) or "too abstract" ($P_{10}$). For example, $P_{10}$ said that "there was truth in there, but it was like a bit drowned out by the things that surrounded it". And $P_{11}$ said that "It just tells me I have to do this and it doesn't come up with the specific examples". $P_{12}$ wished for a "follow-up feature where you can […] ask it again, for example, just something specific".

Closely related, participants perceived the feedback's pattern of connecting the suggestions to the persona's definition as repetitive. For example, $P_9$ said that the first parts "are a description […] But the thing is, I wrote that. I know it's there. […] I want just the last third". Similarly, $P_{10}$ said: "I don't need them to tell me every time who they are." This perception is also reflected in the Likert ratings on suitable length (Figure 8).

As a reminder, our rationale for prompting the system to respond in this way was to reveal to participants how their defined attributes informed the feedback. We return to potential improvements in the discussion (Section 6.5.2).

*5.4.3 Controlling the Model.* Partly in response to the above deficits, several participants brought up ways of steering feedback through additional parameters that users set in the UI. $P_6$ said that "It would have been nice, like, if I could have kind of tuned the feedback I wanted. […] I could have, like, highlighted it, but then also said, like, what type of feedback I want". Similarly, to make feedback more specific, $P_9$ proposed "a third step", where they can specify

"feedback based on the structure or based on the content". Further feature suggestions included that the system should know about previous feedback, as well as giving personas instructions.

## 6 DISCUSSION

### 6.1 Taxonomy of AI Personas for Feedback

Based on our study, we provide a taxonomy of personas and which dimension users considered. The most frequently observed attribute that users took into account was **expertise**. In the specific academic setting we investigated, this was frequently related to competencies, professional roles and qualifications. Such expertise was also defined by mentioning specific, well-known institutions like Oxford University ($P_{10}$) or the New York Times ($P_{13}$). $P_{15}$ used the abstract term "scientific writer", while $P_{11}$ referred to a top researcher by name.

Moreover, **social relation** was an important consideration. This is reflected in personas named "best friend", "my colleague", and "my manager". One participant even mentioned that it might be fun to define a persona such as "mom" ($P_7$). We also observed that the **valence** of personas was taken into account. Some users considered a "worst-case" person like a "mean reviewer" ($P_{12}$), others emulated their "PhD advisor" ($P_{14}$) or their "PhD colleague" ($P_{11}$), as more positively supportive personas. Another dimension we identified was the **level of involvement**, again ranging from closely related personas, such as "my manager" ($P_7$) and "my PhD colleague" ($P_{11}$), to less involved personas such as a "New York Times journalist" ($P_{13}$).

The participants also specified certain characteristics of the personas, e.g., that the grading style of a professor is "strict" ($P_{15}$) or that reviewer's comments are "critical" or "mean" ($P_{12}$). In line with the kind of personas commonly used in design, some participants also mentioned the hobbies of a persona ($P_{11}$).

Our study sample was mostly young academics. We expect that other populations will define different concrete personas. While the presented taxonomy is on a higher level of abstraction, insights from other user groups might extend or refine it. Future work could explore this further with different samples. The taxonomy may also have to be adapted for other writing styles. For instance, expertise may play a less important role in the context of fiction writing. Future work could explore how the different factors must be weighted based on genre.

### 6.2 The Benefits of On-Demand Feedback from AI Personas

We identified three main aspects about AI personas that writers found valuable in our study.

*6.2.1 Writers value and use feedback from AI personas.* The AI personas were useful for our participants. They were viewed as a "cool concept" ($P_7$) and "super helpful" ($P_6$) and recognized as providing different perspectives and helpful feedback. It is particular noteworthy that most participants either kept the feedback and the revised text for their projects beyond the study ($P_9$, $P_{10}$, $P_{11}$, $P_{12}$), or said they would have done so, if the study had been in line with their overall process for that work ($P_7$, $P_8$: if they had more time; and $P_{13}$, $P_{14}$, $P_{15}$, $P_{16}$ had used past/published writing projects). That is, even the feedback provided in the study will already have

real-world implications. In summary, we conclude that the concept and implementation were positively received.

*6.2.2 Writers use AI personas to consider multiple perspectives.* One key benefit that we identified is that AI personas in our prototype empowered users to compare *multiple* perspectives to reflect on their writing (cf. goal 4 in Section 3.1.2). $P_6$ reported several occasions where adding different perspectives made him "step outside of [their] writing and look at it", which helped them notice things they had not noticed before. $P_9$ compared the AI persona approach to the video game Sims, where the user can "create a lot of characters". Strategically, writers added further personas throughout their writing time in the study (Figure 6). Overall, participants created a diverse range of personas (Section 5.3.1, Appendix A.4).

*6.2.3 Writers use AI personas to gain access to perspectives on demand.* The AI personas provided users with *on-demand* feedback from the perspectives of people that they otherwise would not have been able to receive feedback from (cf. goal 1 in Section 3.1.2). Participants liked the convenience of getting "instant feedback" ($P_{14}$). $P_{13}$ also appreciated getting "feedback from these different kinds of people without actually having to find them". $P_7$ described how they could create an avatar of professors that they have not met yet, based on a description of their backgrounds and research interests. While we rarely observed such advanced use of AI personas to "simulate" people a user has not yet met, it shows the potential of the concept and provides ample opportunities for future work.

For this discussion, we reflect on three specific examples of such on-demand feedback: The first example is $P_{14}$, who requested feedback from a professor who taught a particular course. This professor may have moved to a different university, and receiving feedback from them may not be possible anymore. Moreover, due to the many obligations of a professor, they may not have been able to provide feedback on demand. A second example is provided by $P_{16}$, who used the AI persona to receive feedback from a doctor in radiology. Radiologists are highly paid since they are scarce, in high demand, and their services are essential. The tool provided a radiologist's perspective on demand. The third example is $P_{12}$, who used the tool to receive feedback from "the mean reviewer" to improve the text. Reviewing scientific papers is a complex process that requires careful consideration. Accordingly, reviewers require sufficient time to write thoughtful reviews. With the AI personas, $P_{12}$ received feedback from a reviewer perspective already during their work on the manuscript, before sending the paper to actual reviewers.

## 6.3 The Challenges of Working with AI Personas for Feedback

We identified two key challenges: Defining personas and verbose feedback.

*6.3.1 Writers find it difficult to define personas.* Initially defining a persona was challenging because it was difficult for participants to articulate the key aspects of the target group, even with the improved UI in study 2 (see Section 3.1.1 and goal 2 in Section 3.1.2). Nevertheless, they found several strategies, such as thinking of a specific (real) person, using our provided example attributes (Figure 2), and iteratively refining a persona after reading its feedback

(cf. goal 3 in Section 3.1.2). Future research should explore further ways of supporting the creation of AI personas. For example, one idea is to provide a library of personas that allow writers to collaboratively refine and reuse personas. This is akin to prompt collections for LLM-based systems, where users share the instructions that they empirically found to be particularly helpful for a given task. We discuss another direction in Section 6.5.3.

*6.3.2 Generated feedback can be verbose, repetitive and unspecific.* The other key issue that we identified concerns text qualities. Participants found the feedback too verbose and – connected to this – repetitive and unspecific, although our UI enabled them to request feedback for specific text passages, not the whole text (cf. goal 5 in Section 3.1.2). One reason for this was that feedback started by repeating persona information. We discuss further prompt engineering as a solution strategy in Section 6.5.2. Conceptually, based on participants' comments and iterative refinement of personas, we expect tradeoffs between feedback length and helpfulness as well as the transparency of the persona's "argument" (feedback is *this* because of *that* persona information). Future work could use our insights as a starting point to explore the "sweet spot" here. This might also be dynamic. For example, as some comments related to repetitiveness suggest, it might be useful to reduce feedback length over time as users become familiar with a persona.

## 6.4 The Representativity of Contemporary AI Systems

As AI personas are a combination of user-defined personas and LLMs, they have the potential to reproduce harmful stereotypes. In our in-depth qualitative analysis, we did not identify any instances where participants perceived the feedback as perpetuating stereotypes. However, we identified situations where they criticized that the system unnecessarily repeated certain attributes. While seeing how the LLM links personal attributes and feedback in its written output might be insightful (at least initially), it is also a likely place where stereotypes could surface.

Even though we did not encounter stereotypes in practice, the potential dangers of LLMs in this regard have been described in prior work, for instance, by Bender et al. [5]. Prior work provides a broad overview of stereotypical and derogatory associations that LLMs infer from their training data, including sensitive attributes like gender, race, ethnicity, and disability status. Bender et al. argue that reliance on internet data is associated with a risk of encoding the dominant view and failing to represent changing social views after training ("value-lock"). They also explain why the mere size of the training corpus does not guarantee diversity since only a certain privileged group can fully participate online. This connects to work that discussed which human values are represented by LLMs and which ones are not [32].

These concerns are important for writing support systems since prior work showed that predictive keyboards [4] and biased language models [30] can influence people's text and attitudes. While not a major concern in our investigation, we found that $P_6$ "felt pressured to actually make the changes", which indicates the potential power of feedback from AI personas in this regard.

In view of these aspects, we consider it essential that the concept of AI personas, especially when put into practice, is pursued with

careful consideration of how the resulting representations of people can be facilitated and evaluated. Our analysis of the generated feedback did not reveal derogatory or stereotypical statements, such as sexist or racist ones (cf. Section 5.1.3). Nevertheless, systems that provide AI personas could warn users of the potentially limited representativity of contemporary AI systems and how they can arise from limitations in the training data and the training processes.

One key part of our concept is that writers decide on what attributes to define for their personas, which they can change anytime. While that is not the same as having control over the underlying LLM, it provides one pathway of influence to the user, which might be helpful in cases where writers themselves recognize generated output as unsuitable.

That said, there might also be cases where the model's bias is more subtly expressed in the feedback, such that writers may not immediately notice this. Suppose stereotypes like a sexist statement are detected in generated comments in a research setting like our study. In that case, this may be addressed by sending an educational debrief to participants, as done in prior work [30] (e.g., informing participants about the incorrect or biased statement, with a link/-content from a credible source to counter it). In production systems, regular audits may be necessary to ensure that systems are as free of harmful stereotypes as possible [50]. For this, it may be helpful to inform users what stereotypes were detected, how they were detected, and how they were addressed.

Looking ahead, we anticipate that these questions will not remain limited to the design of specific tools but likely impact HCI research more broadly. For example, it seems increasingly likely to receive AI-generated responses when conducting survey-based user research online [22].

## 6.5 Implications for Research and Design

Here we discuss implications for a shift in the design perspective of AI writing tools and two ideas for addressing the identified challenges in the future.

*6.5.1 Shifting the design focus: From human-AI to writer-reader relationships.* Recent related work is often framed around human-AI interaction (e.g. [15, 41, 63]). This emphasizes the relationship of humans and AI (or NLP [62]), where the human in focus of the design and research questions is the writer (cf. [12]). We propose an alternative framing – designing for the *writer-reader relationship*, which AI might support. This expanded view acknowledges two impacted human stakeholders explicitly.

This reframing has implications for research and design, as we demonstrated in this paper, by taking the writer-reader relationship as a starting point for our concept development. Instead of asking how people might write with AI, we asked how AI might support people in writing for people.

Related work on explainable AI (XAI) revealed interesting perspectives by recognizing XAI as "socially-situated" [16]. In our context, our socio-technical expansion of the design perspective on AI writing tools complements recent work on the social dynamics of writing support [19], as well as calls for such tools to consider social aspects of language [38]. It also contributes balance to design perspectives derived from the writing model by Flower and Hayes [17], which is increasingly considered in the design of

writing tools (see [7, 18]) but focuses on processes within the writer and thus has little to say by itself about the perspectives of other affected people, such as readers. Finally, we expect this framing to be fruitful beyond writing, for example, to also keep in mind the creator-audience relationship when designing tools for creators of images, videos, podcasts or games.

*6.5.2 Prompt engineering.* The presented findings are based on one model and one way of prompting it. Thus, we do not claim to present the "optimal" implementation of AI personas for feedback. While further design iterations were beyond the scope of this paper, we have started to explore two strategies to address verbosity, following the insights of our study: One strategy is to modify the prompt template to emphasize more strongly that feedback should be concise. In addition, we shortened the text in the few-shot examples. Anecdotally, first observations indicate that this may not be enough to keep generated comments concise consistently. As another strategy, we experimented with applying the model a second time to its own output, using a second prompt that instructs it to make the given feedback more concise. This seemed promising in our first observations and is in line with "chaining" in other prompt-based systems (cf. [61]).

Future work could explore in detail how such prompting techniques (e.g. [10, 60]) can be leveraged specifically to make LLM output less verbose, since this is a broader issue beyond our use case: For example, Kabir et al. [33] found LLMs to provide verbose answers to software engineering queries, and Sun et al. [58] diagnosed that LLMs struggle with hard constraints, including on length ("numerical planning"). For our use case, further prompts might also be explored with regard to how representative the feedback is (see Section 6.4).

Underlying these considerations is the emerging question in HCI of how to support end users in prompting effectively [65]. Here, the persona concept could be explored as a framing that allows non-experts to steer a model towards a desired role through an understandable UI and metaphor. In a way, this makes explicit that there are (customizable) UI concepts to design around the "system instruction" prompts of systems like ChatGPT (cf. Section 2.2). This direction also connects to recent work on configurable UIs for LLMs [35, 36].

More broadly, with prompt engineering and LLMs, building functional prototypes of AI tools is becoming increasingly feasible in user-centred design. In this light, we see our work as a starting point towards understanding AI personas in various use cases. We hope others will join us to make them even more helpful for writers and other creators.

*6.5.3 Connecting writers and readers through reader-defined personas.* Given the complexity of defining a persona, it is essential to think of ways of supporting this further. One direction could be *reader-defined personas*. Rather than asking individual writers to imagine personas that represent their readers, future work could ask readers to define their personas themselves. This would not only simplify the workflow for writers. Following the mantra "Nothing About Us Without Us", that scholars of Critical Disability Studies in HCI subscribe to [55], it would allow readers to articulate what they need and explicate what the AI-based persona should pay attention to. Conceptually, the AI would then serve both stakeholders

(readers and writers) as a mediating "executable proxy" for communication: For example, the writer could refer to the AI persona in moments where direct communication with the readers is not possible, or use it to get feedback more frequently than what is otherwise feasible. Future work could explore this as an important kind of computer-supported cooperative work that socio-technical interventions can facilitate.

## 7 CONCLUSION

We have proposed the concept of *AI personas for feedback*: Writers define personas of their target readers to receive feedback on their text. This feedback is generated by an LLM based on a prompt that includes the persona information.

In our study, participants used these personas to get feedback from different perspectives but found it challenging to initially define them. The feedback was seen as helpful but too long. It inspired writers to make revisions that benefited the text in their view.

We currently see a "Cambrian explosion" of LLM-based tools. Many use AI to generate text in lieu of human writing and lack a clear use case for doing so. Here, we challenge predominant design assumptions and goals by expanding the socio-technical perspective in the design of AI writing tools. We call for recognizing the social nature of writing as a starting point for tool design. Complementary, we pursued a design goal that builds on the writer-reader relationship.

While AI feedback comes with limitations (e.g. lack of specificity, cf. [19]), it offers enticing benefits, such as direct editor integration, almost instant turnarounds, and infinite availability and patience with a struggling writer's repeated requests. We look forward to future work towards the larger vision of leveraging AI to make us write better, to consider diverse reader perspectives, and to achieve a writing goal that is inherently human: sharing ideas, through language, that others can understand.

To facilitate such future work, we release our prototype and material here:

https://osf.io/4awsm/

## ACKNOWLEDGMENTS

## REFERENCES

[1] 2020. How to Set Goals in the Grammarly Editor, and Why You Should | Grammarly Spotlight. https://www.grammarly.com/blog/set-goals-grammarly-editor/
[2] 2023. MeMo:KI - Opinion Monitor Artificial Intelligence. https://www.cais-research.de/en/research/memoki/
[3] The Team at AMI. 2020. Precursors to a Digital Muse. https://medium.com/artists-and-machine-intelligence/precursors-to-a-digital-muse-2653a3025700
[4] Kenneth C. Arnold, Krysta Chauncey, and Krzysztof Z. Gajos. 2020. Predictive Text Encourages Predictable Writing. In *Proceedings of the 25th International Conference on Intelligent User Interfaces* (Cagliari, Italy) *(IUI '20)*. Association for Computing Machinery, New York, NY, USA, 128–138. https://doi.org/10.1145/3377325.3377523
[5] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 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* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 610–623. https://doi.org/10.1145/3442188.3445922
[6] Michael S. Bernstein, Greg Little, Robert C. Miller, Björn Hartmann, Mark S. Ackerman, David R. Karger, David Crowell, and Katrina Panovich. 2010. Soylent: A Word Processor with a Crowd Inside. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology* (New York, New York, USA) *(UIST '10)*. Association for Computing Machinery, New York, NY, USA, 313–322. https://doi.org/10.1145/1866029.1866078
[7] Advait Bhat, Saaket Agashe, Parth Oberoi, Niharika Mohile, Ravi Jangir, and Anirudha Joshi. 2023. Interacting with Next-Phrase Suggestions: How Suggestion Systems Aid and Influence the Cognitive Processes of Writing. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 436–452. https://doi.org/10.1145/3581641.3584060
[8] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (jan 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa
[9] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (2019), 589–597. https://doi.org/10.1080/2159676X.2019.1628806 arXiv:https://doi.org/10.1080/2159676X.2019.1628806
[10] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 1877–1901. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf
[11] Daniel Buschek, Malin Eiband, and Heinrich Hussmann. 2022. How to Support Users in Understanding Intelligent Systems? An Analysis and Conceptual Framework of User Questions Considering User Mindsets, Involvement, and Knowledge Outcomes. *ACM Trans. Interact. Intell. Syst.* 12, 4, Article 29 (nov 2022), 27 pages. https://doi.org/10.1145/3519264
[12] Minsuk Chang, John Joon Young Chung, Katy Ilonka Gero, Ting-Hao Kenneth Huang, Dongyeop Kang, Mina Lee, Vipul Raheja, and Thiemo Wambsganss. 2023. The Second Workshop on Intelligent and Interactive Writing Assistants. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 369, 5 pages. https://doi.org/10.1145/3544549.3573826
[13] Alan Cooper, Robert Reimann, David Cronin, and Christopher Noessel. 2014. *About Face: The Essentials of Interaction Design* (4 ed.). Wiley, Indianapolis, IN.
[14] Juliet Corbin and Anselm Strauss. 2014. *Basics of qualitative research: Techniques and procedures for developing grounded theory.* Sage publications.
[15] Hai Dang, Karim Benharrak, Florian Lehmann, and Daniel Buschek. 2022. Beyond Text Generation: Supporting Writers with Continuous Automatic Text Summaries. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 98, 13 pages. https://doi.org/10.1145/3526113.3545672
[16] Upol Ehsan, Q. Vera Liao, Michael Muller, Mark O. Riedl, and Justin D. Weisz. 2021. Expanding Explainability: Towards Social Transparency in AI systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 82, 19 pages. https://doi.org/10.1145/3411764.3445188
[17] Linda Flower and John R Hayes. 1981. A cognitive process theory of writing. *College composition and communication* 32, 4 (1981), 365–387.
[18] Katy Gero, Alex Calderwood, Charlotte Li, and Lydia Chilton. 2022. A Design Space for Writing Support Tools Using a Cognitive Process Model of Writing. In *Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022)*. Association for Computational Linguistics, Dublin, Ireland, 11–24. https://doi.org/10.18653/v1/2022.in2writing-1.2
[19] Katy Ilonka Gero, Tao Long, and Lydia B Chilton. 2023. Social Dynamics of AI Support in Creative Writing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 245, 15 pages. https://doi.org/10.1145/3544548.3580782
[20] Susanne Göpferich. 2009. Comprehensibility assessment using the Karlsruhe comprehensibility concept. *The Journal of Specialised Translation* 11, 2009 (2009), 31–52.
[21] Emanuela Haller and Traian Rebedea. 2013. Designing a Chat-bot that Simulates an Historical Figure. In *2013 19th International Conference on Control Systems and Computer Science*. 582–589. https://doi.org/10.1109/CSCS.2013.85
[22] Perttu Hämäläinen, Mikke Tavast, and Anton Kunnari. 2023. Evaluating Large Language Models in Generating Synthetic HCI Research Data: A Case Study. In

*Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 433, 19 pages. https://doi.org/10.1145/3544548.3580688

[23] Jeffrey T Hancock, Mor Naaman, and Karen Levy. 2020. AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations. *Journal of Computer-Mediated Communication* 25, 1 (01 2020), 89–100. https://doi.org/10.1093/jcmc/zmz022 arXiv:https://academic.oup.com/jcmc/article-pdf/25/1/89/32961176/zmz022.pdf

[24] John Hayes and Linda Flower. 1980. *Identifying the organization of writing processes.*

[25] John R. Hayes. 2012. Modeling and Remodeling Writing. *Written Communication* 29, 3 (July 2012), 369–388. https://doi.org/10.1177/0741088312451260 Publisher: SAGE Publications Inc.

[26] Hendrik Heuer and Elena Leah Glassman. 2023. Accessible Text Tools for People with Cognitive Impairments and Non-Native Readers: Challenges and Opportunities. In *Proceedings of Mensch Und Computer 2023* (Rapperswil, Switzerland) *(MuC '23)*. Association for Computing Machinery, New York, NY, USA, 250–266. https://doi.org/10.1145/3603555.3603569

[27] Hendrik Heuer and Elena Leah Glassman. 2023. Accessible Text Tools: Where They Are Needed & What They Should Look Like. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 20, 7 pages. https://doi.org/10.1145/3544549.3585749

[28] Jess Hohenstein and Malte Jung. 2020. AI as a moral crumple zone: The effects of AI-mediated communication on attribution and trust. *Computers in Human Behavior* 106 (2020), 106190. https://doi.org/10.1016/j.chb.2019.106190

[29] Ting-Hao 'Kenneth' Huang, Vipul Raheja, Dongyeop Kang, John Joon Young Chung, Daniel Gissin, Mina Lee, and Katy Ilonka Gero (Eds.). 2022. *Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022)*. Association for Computational Linguistics, Dublin, Ireland. https://aclanthology.org/2022.in2writing-1.0

[30] Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. 2023. Co-Writing with Opinionated Language Models Affects Users' Views. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 111, 15 pages. https://doi.org/10.1145/3544548.3581196

[31] Maurice Jakesch, Megan French, Xiao Ma, Jeffrey T. Hancock, and Mor Naaman. 2019. AI-Mediated Communication: How the Perception That Profile Text Was Written by AI Affects Trustworthiness. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300469

[32] Rebecca L Johnson, Giada Pistilli, Natalia Menédez-González, Leslye Denisse Dias Duran, Enrico Panai, Julija Kalpokiene, and Donald Jay Bertulfo. 2022. The Ghost in the Machine has an American accent: value conflict in GPT-3. *arXiv preprint arXiv:2203.07785* (2022).

[33] Samia Kabir, David N Udo-Imeh, Bonan Kou, and Tianyi Zhang. 2023. Who Answers It Better? An In-Depth Analysis of ChatGPT and Stack Overflow Answers to Software Engineering Questions. *arXiv preprint arXiv:2308.02312* (2023).

[34] Judy Kay and Bob Kummerfeld. 2013. Creating Personalized Systems That People Can Scrutinize and Control: Drivers, Principles and Experience. *ACM Trans. Interact. Intell. Syst.* 2, 4, Article 24 (jan 2013), 42 pages. https://doi.org/10.1145/2395123.2395129

[35] Tae Soo Kim, Yoonjoo Lee, Minsuk Chang, and Juho Kim. 2023. Cells, Generators, and Lenses: Design Framework for Object-Oriented Interaction with Large Language Models. In *The 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23), October 29-November 1, 2023, San Francisco, CA, USA* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, 18 pages. https://doi.org/979-8-4007-0132-0/23/10

[36] Tae Soo Kim, Arghya Sarkar, Yoonjoo Lee, Minsuk Chang, and Juho Kim. 2023. LMCanvas: Object-Oriented Interaction to Personalize Large Language Model-Powered Writing Environments. http://arxiv.org/abs/2303.15125 arXiv:2303.15125 [cs].

[37] Matthew G. Kirschenbaum. 2016. *Track Changes: A Literary History of Word Processing*. Harvard University Press.

[38] Vivek Kulkarni and Vipul Raheja. 2023. Writing Assistants Should Model Social Factors of Language. https://doi.org/10.48550/arXiv.2303.16275 arXiv:2303.16275 [cs].

[39] Inghard Langer, Friedemann Schulz von Thun, and Reinhard Tausch. 2019. *Sich verständlich ausdrücken*. Ernst Reinhardt Verlag.

[40] Mina Lee, Katy Ilonka Gero, John Joon Young Chung, Simon Buckingham Shum, Vipul Raheja, Hua Shen, Subhashini Venugopalan, Thiemo Wambsganss, David Zhou, Emad A. Alghamdi, Tal August, Avinash Bhat, Madiha Zahrah Choksi, Senjuti Dutta, Jin L.C. Guo, Md Naimul Hoque, Yewon Kim, Simon Knight, Seyed Parsa Neshaei, Antonette Shibani, Disha Shrivastava, Lila Shroff, Agnia Sergeyuk, Jessi Stark, Sarah Sterman, Sitong Wang, Antoine Bosselut, Daniel Buschek, Joseph Chee Chang, Sherol Chen, Max Kreminski, Joonsuk Park, Roy Pea, Eugenia Ha Rim Rho, Zejiang Shen, and Pao Siangliulue. 2024. A Design Space for Intelligent and Interactive Writing Assistants. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA.

[41] Mina Lee, Percy Liang, and Qian Yang. 2022. CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Capabilities. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI'22)*. Association for Computing Machinery, New Orleans, LA, USA. https://doi.org/10.1145/3491102.3502030

[42] Nina Macdonald, Lawrence Frase, P Gingrich, and Stacey Keenan. 1982. The writer's workbench: Computer aids for text analysis. *IEEE Transactions on Communications* 30, 1 (1982), 105–110.

[43] Philipp Mayring. 2021. Qualitative content analysis: A step-by-step guide. *Qualitative Content Analysis* (2021), 1–100.

[44] Hannah Mieczkowski, Jeffrey T. Hancock, Mor Naaman, Malte Jung, and Jess Hohenstein. 2021. AI-Mediated Communication: Language Use and Interpersonal Effects in a Referential Communication Task. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 17 (apr 2021), 14 pages. https://doi.org/10.1145/3449091

[45] David Nilsson, Magnus Sahlgren, and Jussi Karlgren. 2016. Dead Man Tweeting. In *Workshop on Collecting and Generating Resources for Chatbots and Conversational Agents-Development and Evaluation, Portorož (Slovenia), May 28, 2016.*

[46] Joon Sung Park, Joseph C O'Brien, Carrie J Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. *arXiv preprint arXiv:2304.03442* (2023).

[47] Joon Sung Park, Lindsay Popowski, Carrie Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2022. Social Simulacra: Creating Populated Prototypes for Social Computing Systems. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 74, 18 pages. https://doi.org/10.1145/3526113.3545616

[48] Peter Rechenberg. 2006. Technisches Schreiben. In *Technisches Schreiben*. Carl Hanser Verlag GmbH & Co. KG, 1–10. https://doi.org/10.3139/9783446409576.fm

[49] Kevin Roose. 2022. The brilliance and weirdness of ChatGPT. https://www.nytimes.com/2022/12/05/technology/chatgpt-ai-twitter.html. *The New York Times* (2022).

[50] Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014. Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and discrimination: converting critical concerns into productive inquiry* 22, 2014 (2014), 4349–4357.

[51] Oliver Schmitt and Daniel Buschek. 2021. CharacterChat: Supporting the Creation of Fictional Characters through Conversation and Progressive Manifestation with a Chatbot. In *Creativity and Cognition* (Virtual Event, Italy) *(C&C '21)*. Association for Computing Machinery, New York, NY, USA, Article 10, 10 pages. https://doi.org/10.1145/3450741.3465253

[52] Haoyu Song, Yan Wang, Kaiyan Zhang, Wei-Nan Zhang, and Ting Liu. 2021. BoB: BERT Over BERT for Training Persona-based Dialogue Models from Limited Personalized Data. https://doi.org/10.48550/arXiv.2106.06169 arXiv:2106.06169 [cs].

[53] Haoyu Song, Wei-Nan Zhang, Yiming Cui, Dong Wang, and Ting Liu. 2019. Exploiting Persona Information for Diverse Generation of Conversational Responses. https://doi.org/10.48550/arXiv.1905.12188 arXiv:1905.12188 [cs].

[54] Jared Spataro. 2023. Introducing Microsoft 365 Copilot – your copilot for work. https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/

[55] Katta Spiel, Kathrin Gerling, Cynthia L. Bennett, Emeline Brulé, Rua M. Williams, Jennifer Rode, and Jennifer Mankoff. 2020. Nothing About Us Without Us: Investigating the Role of Critical Disability Studies in HCI. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3375150

[56] Carola Strobl, Emilie Ailhaud, Kalliopi Benetos, Ann Devitt, Otto Kruse, Antje Proske, and Christian Rapp. 2019. Digital support for academic writing: A review of technologies and pedagogies. *Computers & Education* 131 (2019), 33–48. https://doi.org/10.1016/j.compedu.2018.12.005

[57] Afra Sturm, Mirjam Weder, and Mirjam Weder. 2016. *Schreibkompetenz, Schreibmotivation, Schreibförderung: Grundlagen und Modelle zum Schreiben als soziale Praxis*. Klett Kallmeyer.

[58] Jiao Sun, Yufei Tian, Wangchunshu Zhou, Nan Xu, Qian Hu, Rahul Gupta, John Frederick Wieting, Nanyun Peng, and Xuezhe Ma. 2023. Evaluating Large Language Models on Controlled Generation Tasks. arXiv:2310.14542 [cs.CL]

[59] Yuqian Sun, Xingyu Li, and Ze Gao. 2023. Inspire creativity with ORIBA: Transform Artists' Original Characters into Chatbots through Large Language Model. arXiv:2306.09776 [cs.MM]

[60] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems* 35 (2022), 24824–24837.

[61] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language

Model Prompts. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>New Orleans</city>, <state>LA</state>, <country>USA</country>, </conf-loc>) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 385, 22 pages. https://doi.org/10.1145/3491102.3517582

[62] Qian Yang, Justin Cranshaw, Saleema Amershi, Shamsi T. Iqbal, and Jaime Teevan. 2019. Sketching NLP: A Case Study of Exploring the Right Things To Design with Language Intelligence. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300415

[63] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-Examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376301

[64] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: Story Writing With Large Language Models. In *27th International Conference on Intelligent User Interfaces (IUI '22)*. Association for Computing Machinery, New York, NY, USA, 841–852. https://doi.org/10.1145/3490099.3511105

[65] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 437, 21 pages. https://doi.org/10.1145/3544548.3581388

[66] Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing dialogue agents: I have a dog, do you have pets too? *arXiv preprint arXiv:1801.07243* (2018).

[67] Yinhe Zheng, Rongsheng Zhang, Minlie Huang, and Xiaoxi Mao. 2020. A Pre-Training Based Personalized Dialogue Generation Model with Persona-Sparse Data. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 05 (April 2020), 9693–9700. https://doi.org/10.1609/aaai.v34i05.6518 Number: 05.

## A APPENDIX

## A.1 Prompt Template for Generating Feedback

This is the prompt template in Python that we used (some formatting added for the paper). The few-shot examples also followed this template and were inserted between the system description and the final instruction, as indicated below. This format is expected by the OpenAI API (Chat Completion).

```
{
    "role": "system",
    "content": 'Personas are defined using four fixed attributes:
role, background, style, and content. Each attribute consists of
user-defined key-value pairs. The possible key-value pairs are not
 predefined and can vary. Generate persona-specific feedback for
the text snippet highlighted by the user, considering the persona'
s unique attributes and any additional key-value pairs that might
be defined by the user. You will take the role of the persona and
write from their viewpoint. Every key-value attribute that is
included in the personas definition describes the persona and
therefore you. The feedback should align with the persona's
characteristics and viewpoint, providing insights, suggestions, or
 comments that are relevant to the persona\'s role, background,
style preferences, and content preferences.

    Input:
    Text: "Selected text snippet from the user\'s editor."
    Persona:
    - Role: {"key": "value"}
    - Background: {"key": "value"}
    - Style: {"key": "value"}
    - Content: {"key": "value"}
```

```
    Output: Generate persona-specific feedback for the provided
text snippet based on the given persona attributes. Write the
feedback as if you would be this persona. Consider the role,
background, style, and content preferences of the persona. Provide
 insights, suggestions, or comments that align with the persona\'s
 characteristics and viewpoint. Feel free to incorporate any
additional key-value pairs defined by the user in the persona
definition to enhance the relevance of the feedback. Write one
continuous feedback that is not longer than 200 words.',
},
...few-shot examples...
{
    "role": "user",
    "content": 'Input:
    Text: "'
    + selectedText
    + '"
    Persona:
    - Role: '
    + str(roleAttributes)
    + "
    - Background: "
    + str(backgroundAttributes)
    + "
    - Style: "
    + str(styleAttributes)
    + "
    - Content: "
    + str(contentAttributes),
}
```

The attribute-description pairs defined by the writers were inserted as indicated by the Python statements above. For example, if a participant defined a persona with the pairs "role: reviewer", "writing style: formal", "sentence length: short", and "occupation: CS professor", and selected a text "Lorem ipsum dolor sit amet", then the corresponding part of the prompt template above would concretely look like this:

```
    Input:
    Text: "Lorem ipsum dolor sit amet"
    Persona:
    - Role: {"role": "reviewer"}
    - Background: {"occupation": "CS professor"}
    - Style: {"writing style": "formal", "sentence length": "short
"}
    - Content: {}
```

## A.2 Feedback Interaction Example

Figure 7 shows an example of one concrete moment of interaction of one participant ($P_{14}$) with two of their AI personas. As shown, the participant selected a paragraph in their draft and subsequently requested feedback from two personas (a reviewer and an Anthropology professor acting as an editor). The participant made concrete edits based on both generated feedback and commented on this experience positively (speech bubbles in the figure).

Note that we replaced the topic of the text in this example to preserve anonymity and avoid publishing the participant's text here. Concretely, we replaced the original topic and text with text from the Wikipedia entry on "human"[3]. The sentence structure, phrasing, semantics/scope of feedback and edits, etc. are all preserved as precisely as possible – we only replaced the topic.
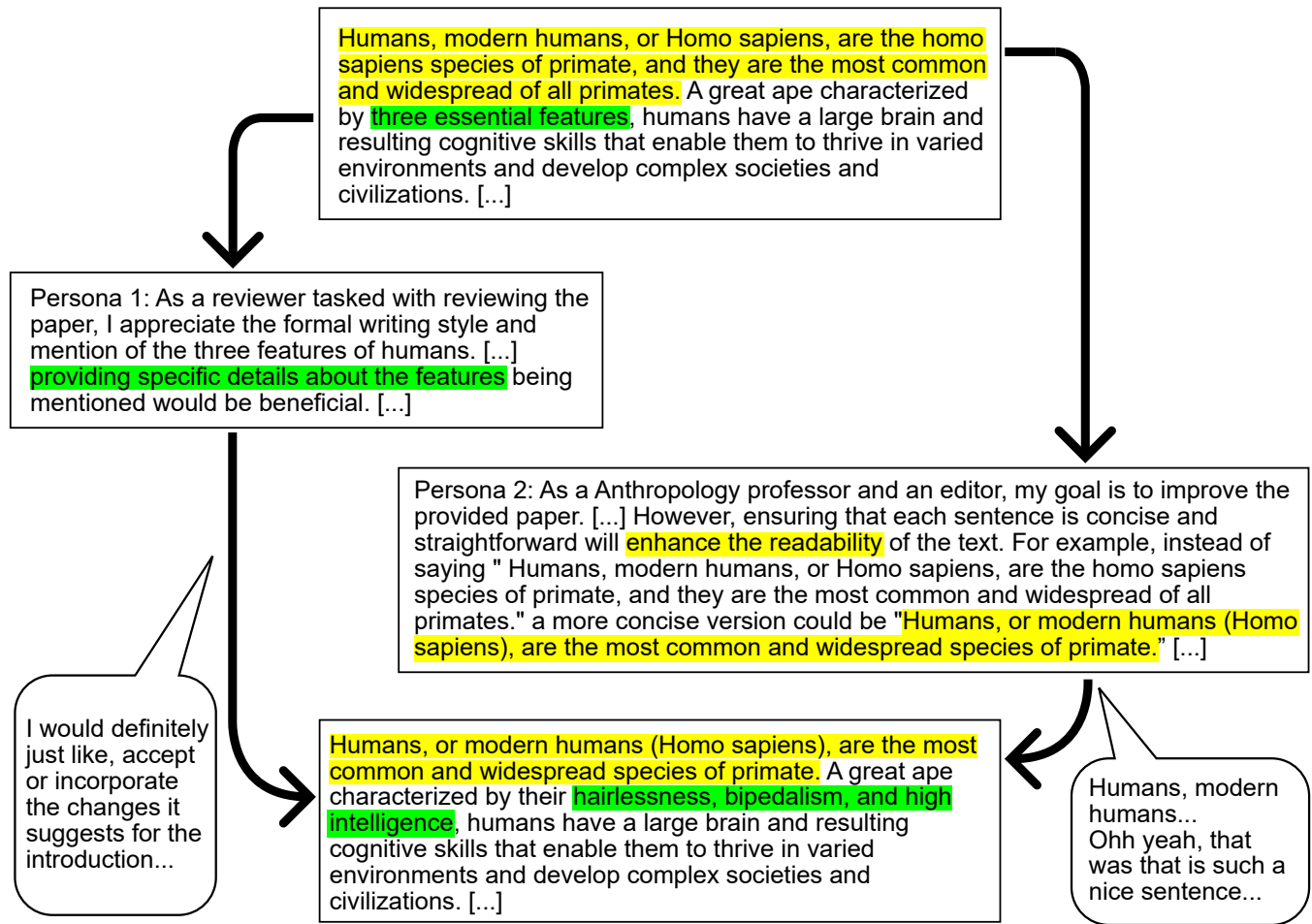
---

[3]https://en.wikipedia.org/wiki/Human

**Figure 7: Example of a moment of interaction of $P_{14}$ with two of their AI personas: $P_{14}$ selected a paragraph in their draft (top box) and requested feedback from two personas (two boxes in the centre). The participant then made concrete edits based on both generated feedback (new draft version in the bottom box), and commented on this experience (speech bubbles). Ellipses ([...]) and color were used for this figure, not a part of how the text was shown in the UI. Note that we replaced the topic of the text in this example to preserve anonymity (see Appendix A.2).**

## A.3 Likert Results

Figure 8 shows the results from the final questionnaire from study 2. The items on interest in AI (bottom 4 items) were taken from the "MeMo:KI – Opinion Monitor AI" [2].

## A.4 Persona Types

Table 2 shows the different personas created in the first study, based on what participants entered for the persona (prototype V1 had a single text box, see Section 3.1.1). For this overview, one researcher coded these and grouped them, resulting in the two shown dimensions (Gender, Age) with an additional "note" column containing further details. Prototype V1 had no dedicated name fields, so the three recorded names reflect how that participant referred to the persona in their verbal comments.

Table 3 shows the different personas created in the second study, where we logged the defined attributes and descriptions. For this overview, one researcher coded these and grouped them, resulting in the five shown dimensions ([Level of] Abstraction, [Writing] Style, Role, Focus [on], Background [related to text topic]) and their levels. While the names are reported as entered by participants, the "note" column displays interesting additional information that was revealed by the participants during the study or in the following interview.

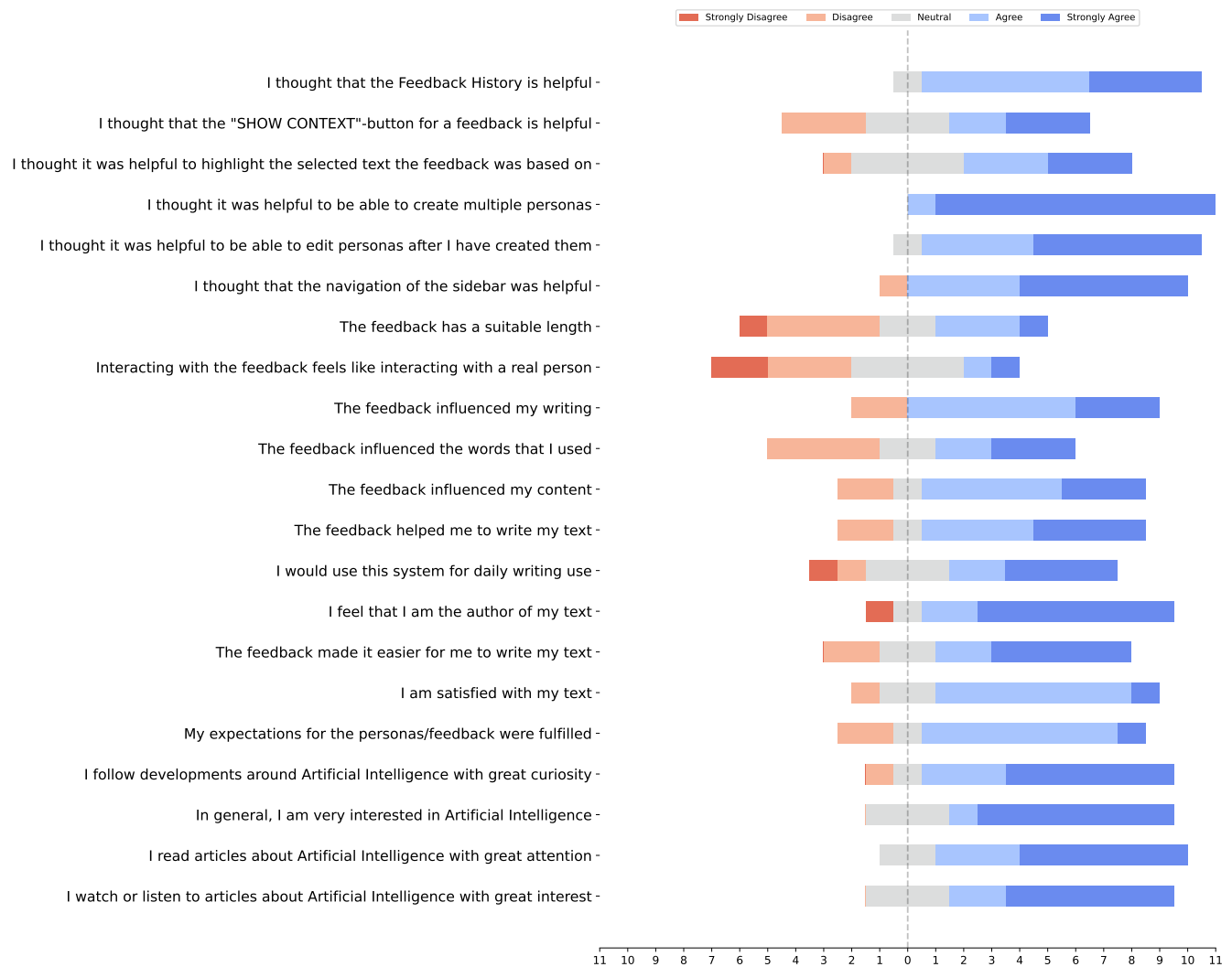**Figure 8: The Likert items on the concept and prototype, as well as interest in AI, asked at the end of study 2. Overall, participants appreciated the provided features. Most found that the feedback was helpful and had an influence on their work. They still perceived themselves as the authors. Participants were critical of the feedback length (too long). Finally, our sample had a high interest in AI.**

| P | Name | Gender | Age | Note |
|---|---|---|---|---|
| 1 | Professor X | male | - | Professor for social gatherings |
| 1 | Frederik | male | 12 | Doesn't like reading |
| 1 | John | male | - | Persona occurs inside the narrative |
| 2 | - | male | 10 | - |
| 2 | - | female | middle aged | housewife |
| 2 | - | - | - | school teacher |
| 2 | - | - | middle aged | worker that wants to use AI |
| 2 | - | - | 60 | unfamiliar with chatbots |
| 2 | - | - | - | teacher that wants to use AI |
| 3 | - | - | - | specialized researcher |
| 3 | - | - | - | specialized reviewer from the US |
| 3 | - | - | - | specialized reviewer from the UK |
| 3 | - | female | - | specialized reviewer from Australia |
| 3 | - | male | - | specialized reviewer from India |
| 3 | - | female | - | specialized reviewer from India |
| 3 | - | male | - | specialized reviewer from China |
| 3 | - | female | - | specialized reviewer from China |
| 3 | - | female | - | specialized reviewer from the UK |
| 3 | - | male | - | specialized reviewer from the UK |
| 4 | - | - | - | cybersecurity researcher specialized |
| 4 | - | - | - | cybersecurity researcher unspecialized |
| 4 | - | - | - | politician with no technical knowledge |
| 4 | - | - | - | government worker for funding research |
| 4 | - | - | - | researcher evaluating novelty |
| 4 | - | - | - | same researcher checking improvements |
| 5 | - | - | - | professor |
| 5 | - | - | - | Italian chef |
| 5 | - | - | - | historian |

**Table 2: Overview of the different personas created in the first study (using prototype V1), based on what participants entered in the single persona description text box that was available in that prototype.**

| P | Persona Name | Note | Abstraction | Style | Role | Focus | Background |
|---|---|---|---|---|---|---|---|
| 6 | Hardware Expert | for fact checking | medium | formal | general | - | related |
| 6 | Writing Editor | for accessibility | high | casual | general | text | unrelated |
| 6 | Average Joe | uninterested perspective | medium | casual | general | text | unrelated |
| 7 | HR Manager Michael | - | real person | formal | specific | facts | related |
| 7 | Sybille | participant's mom | real person | formal | general | - | unrelated |
| 8 | Crypto Bro | perspective of reader | medium | casual | general | - | related |
| 8 | Reviewer | is a professor | high | formal | general | - | related |
| 9 | Persona 1 | is a HCI professor | high | formal | general | - | related |
| 9 | Persona 2 | writes solutions | high | formal | specific | - | related |
| 10 | General Public | educated but not expert | medium | formal | general | text | unrelated |
| 10 | CS Professor, not HCI | Prof. in Computer Science (CS) | low | formal | general | text | unrelated |
| 10 | CS Professor, HCI | Oxford Prof. in CS and HCI | low | formal | general | both | related |
| 11 | [last name] | famous professor | real person | formal | specific | text | related |
| 11 | Persona 2 | for rewriting text | high | formal | specific | facts | related |
| 12 | CHI Reviewer | - | low | formal | general | - | unrelated |
| 12 | Researcher | find good citations | high | - | specific | - | unrelated |
| 13 | PhD Advisor | modelled after real person | real person | formal | specific | facts | related |
| 13 | New to field | NYTimes Journalist for simple text | medium | casual | general | text | unrelated |
| 13 | Negative Reviewer | worst-case persona | medium | - | general | both | related |
| 14 | Persona 1 | - | high | formal | general | both | - |
| 14 | Persona 2 | professor for grading | medium | formal | general | text | related |
| 15 | Scientific Writer | ghostwriter | medium | - | specific | text | unrelated |
| 15 | Scientific Journalist | find specific information | medium | formal | specific | facts | unrelated |
| 15 | Professor | professor for grading | medium | - | general | both | - |
| 16 | Doctor | expert in the field | medium | formal | general | both | related |
| 16 | Best friend | for general support | low | casual | general | - | unrelated |

**Table 3: Overview of the different personas created in the second study (using prototype V2), based on the logged persona attributes and descriptions.**