Pro-human AI System Design: Preserving Human Characteristics in Practice

Rebekka von Wartburg-Kottler
Zurich University of Applied Sciences
VT1 as part of the Master of Science in Engineering
Centre for Artificial Intelligence
Email: vonwareb@students.zhaw.ch

***Abstract*—Artificial intelligence (AI) systems have become an integral part of our lives and influence how we live, work, and act. They interact with us on a fundamental and personal level, touching on characteristics that define us at our core, such as relationality, responsibility, and autonomy. This thesis examines how AI systems can be designed to increase efficiency in a pro-human manner, while preserving these characteristics, summarized as dimensions of the *Humatrix*. The analysis follows a three-step approach. First, it identifies the constitutive human characteristics that must be preserved to remain fully human. Second, using the use case of *Psychiatric Session Reporting*, it examines how manipulative interactions with a particular AI system, the *Base AI System*, can diminish dimensions of the Humatrix by changing the *Lived Expressions* through which these dimensions are enacted in practice. Finally, a concrete design intervention, the *Pro-human Prototype AI System*, is presented to illustrate how the weakening of these dimensions can be reduced. The results suggest that such diminishing arises primarily from routine and habitual use and the associated recurring interactions, rather than from the use of the AI system as such. Assistive text completion is presented as a pro-human system design intervention. This allows the efficiency gains to be maintained while enabling the Lived Expressions through which the affected dimensions of the Humatrix are realized. This approach is not proposed as a definitive solution, but as a guiding orientation for pro-human AI system design.**

***Index Terms*—pro-human, core characteristics, AI systems, system interactions, system design interventions, psychiatric documentation**

1. Introduction

With their remarkable achievements and rapid progress, artificial intelligence (AI) systems have become an integral part of everyday life. They provide support in areas such as communication, decision-making, education, health, and administration (Shneiderman, 2022) and were primarily designed to increase efficiency, predictive accuracy, and optimization, making them powerful tools (Burwell, 2025; Devgun, 2023). The widespread use of such systems shapes the way people live, work, and act, while also changing their perception of the world and the way they make decisions (Ihde, 2010; Verbeek, 2005). At the same time, the increasing performance of these AI systems has fundamentally changed the way people address and solve problems. In the past, goals had to be achieved through independent searching, trial and error, learning from mistakes, and optimization. Today, systems such as ChatGPT (openAI, 2022), Claude (AnthropicPBC, 2025), and Gemini (Google, 2025) significantly shorten this process, as they can easily achieve a goal without the supposedly tedious detours. However, it is precisely these detours and wrong turns that form an essential part of human development and cognitive processes (Dehaene, 2021). Through repeated, embodied, and socially embedded practice, humans acquire skills such as learning, judgment, decision-making, and the ability to act (Vallor, 2016). A recent article in the New York Times illustrates a downside of the rapid access to such AI systems. They stated that OpenAI had modified the design of ChatGPT (openAI, 2022) to increase interaction with their product. According to Hill & Valentino-DeVries, this adjustment led to 50 cases of psychosis, nine hospitalizations, and three deaths (Hill & Valentino-DeVries, 2025). This makes it clear, that AI systems influence not only our work, but also interact with us on a much more fundamental and personal level of human experience. AI systems touch on the core of what defines and constitutes us as human beings, such as, relationality, responsibility, or autonomy. As the article of the New York Times Hill & Valentino-DeVries illustrates, in the worst case, they can negatively manipulate this core. Rather than viewing AI systems exclusively as tools for increasing efficiency, they should be understood as technologies for human self-design (Vallor, 2024a).

This raises the following key question:

How can AI systems be designed to increase efficiency in a pro-human manner, while preserving and ideally even strengthening those characteristics that define and constitute human beings, such as relationality, responsibility, and autonomy?

The primary goal of this thesis is to present an approach for pro-human AI system design from both a methodological and a design-oriented perspective, in response to the key question formulated above. To this end, the proposed approach is structured into three analytical and conceptual steps:

- **Step 1** focuses on identifying the characteristics that constitute the core of being human (henceforth referred to as dimensions of the *Humatrix*). These dimensions make us unique as human beings and should be preserved in order to remain fully human (see Section 2).
- **Step 2** analyzes a specific use case to determine how a particular AI system (henceforth referred to as *Base AI System*)[1] interacts with individual dimensions of the Humatrix and at which points these interactions could potentially lead to manipulative or otherwise negative effects (see Section 3).
- **Step 3** translates the insights gained into a concrete design suggestion that outlines how the Base AI System could be modified toward a pro-human orientation (henceforth referred to as the *Pro-human Prototype AI System*)[1] (see Section 3).

2. Identifying Dimensions of the Humatrix (Step 1)

The question of what define and distinguish humans from other living beings has fascinated humanity for centuries (Gehlen & Rehberg, 2016; Plessner, 1928; Ross, 1956; Scheler, 2016). In light of the rapid developments in AI (Olson, 2025; Rashid & Kausik, 2024; Tegmark, 2018), this question is becoming relevant again in a new form. Such AI systems are beginning to mimic abilities that were once considered uniquely human. They are creative, demonstrate understanding, and can make decisions (Lawrence, 2024). Consequently, these technologies are increasingly influencing human thought, action, and experience and thus, humanity itself. Although earlier thinkers focused mainly on differences between humans and animals (Plessner, 1928; Scheler, 2016), the current debate is increasingly concerned with the relationship between humans and technological innovations and their influence on humanity (Fuchs, 2024; Lawrence, 2024). This section deals with the core characteristics that define and constitute human beings and that must be preserved in order to remain fully human. A systematic weakening of these characteristics would fundamentally impair the essence of being human. On this basis, six dimensions, summarized as the Humatrix were identified, derived from the interdisciplinary fields of philosophy, anthropology, psychology, sociology, and theology (see Figure 1 and Sections 2.1 - 2.6).

1. "Thanks to AlpineAI (AlpineAI, 2026), which is providing its proven and established AI system, *SwissGPT*, for this thesis, thus creating the foundation for demonstrating the implementation of a pro-human design approach."

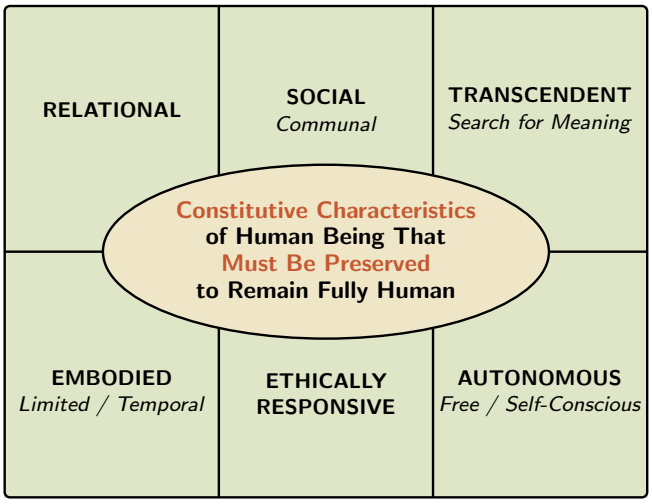


Figure 1: Based on interdisciplinary literature from philosophy, anthropology, psychology, sociology, and theology, six constitutive human characteristics were identified as whose preservation is necessary to remain fully human, summarized as the dimensions of the Humatrix.

2.1. Relational

Human beings do not realize themselves in isolation, but rather through direct encounters with others, grounded in mutual interaction and recognition (Gunkel & Wales, 2021, p. 476-478) [2]. Only within such personal, face-to-face relationships do individuals come to experience themselves as persons (Buber & Lang, 2021) [3]. They are beings whose meaning, identity, and freedom are rooted in relationships (Gastmans et al., 2024, p. 791) [4]; (Calo, 2024, p. 223) [5]. In this sense, human beings are understood as concrete individuals who express their personality in mutual relationships of self-giving and thereby become receptive to others (Gunkel & Wales, 2021, p. 478) [6]. This includes consciously and empathetically empathizing with another person, taking their perspective, understanding their feelings, and responding

2. "The form of divine relationality is that of giving the totality of one's being to another; and so the natural creaturely "person" came to be seen as not a social role but a concrete individual, who exercises his or her personhood in mutual relationships of self-gift that are self-expressive and other-receiving." (Gunkel & Wales, 2021, p. 478)

3. "Der Mensch wird am Du zum Ich." (Buber & Lang, 2021, p. 33)

4. "... indicates that human emotions, such as love, loyalty, fidelity, and empathic concern, as well as the need for affiliation and the need to matter positively to others are the specific features of human existence that are both derived from and contribute to the relationships that make human being human." (Gastmans et al., 2024, p. 791)

5. "... relationality also concerns the ways in which persons express and experience freedom with and through others" (Calo, 2024, p. 223)

6. "The form of divine relationality is that of giving the totality of one's being to another; and so the natural creaturely "person" came to be seen as not a social role but a concrete individual, who exercises his or her personhood in mutual relationships of self-gift that are self-expressive and other-receiving." (Gunkel & Wales, 2021, p. 478)

accordingly (Gunkel & Wales, 2021, p. 479) [7]; (Rubin et al., 2024, p. 2) [8]. The capacity for genuine reciprocity and shared lived experience is therefore a defining feature of relationality and the basis of human existence (Gunkel & Wales, 2021, p. 478) [9]; (Buber & Lang, 2021, p. 16) [10].

2.2. Social *Communal*

Humans are social and community-oriented beings that exist within shared social structures rather than exclusive personal relationships (Fuchs, 2024). This social space consists of shared experiences, narratives, norms, and created symbols (Segessenmann et al., 2025, p. 187) [11]. Unlike the personal relationship of relationality, this refers to the embedding of people in a social structure, with values and spheres of meaning(Gastmans et al., 2024, p. 792) [12]. In these social structures, humans gain orientation, identity, and meaning as members of a collective (Fuchs, 2024; Segessenmann et al., 2025). Human existence is therefore a social existence characterized by shared and jointly responsibility and co-existence within a plurality of others (Segessenmann et al., 2025, p. 187) [13].

2.3. Transcendent *Truth-Seeking / Searching for Meaning / Curious / Self-Reflective*

Humans live not only in a biological sphere, but also in a spiritual dimension, which enables them to think, understand, and question meaning (Scheler, 2016, p. 32) [14]; (Plessner, 1928, p. 309) [15]. In doing so, they make themselves the object of thought and are able to reflect on their own existence (Plessner, 1928). Their thinking and understanding is not purely rational, but is influenced by social and historical contexts (Vallor, 2024b) [16]. This is a lifelong, communal process of interpretation, doubt, and searching (Crawford, 2015; Vallor, 2024b). Humans want to actively shape their world through language, art, religion, and science, and create culture as their own meaning (Cassirer, 1960; Gehlen & Rehberg, 2016).

2.4. Autonomous *Free / Self-Concious*

Being human means shaping one's own life in a self-determined, free, and responsible manner and actively shaping the world from this freedom (Calo, 2024; Segessenmann et al., 2025). True freedom is not independence from others, but inner freedom, grounded in self-consciousness and reflection. This enables humans to define norms, question them critically, and make decisions based on them (Gastmans et al., 2024) [17]. This conscious self-determination leads to a degree of independence that makes humans unique (Calo, 2024; Segessenmann et al., 2025).

2.5. Ethically Responsive

Humans are responsible, moral beings (Gastmans et al., 2024, p. 791) [18]; (Segessenmann et al., 2025). This attitude includes the ability to reflect on one's own actions, assess their consequences, and act responsibly towards oneself and

7. "A person's consciousness is more than what humans seem to share with gorillas; it is a consciousness that voluntarily reaches out to make contact with the consciousness of others as an act of self-giving; it is subjectivity oriented to inter-subjectivity. The mutual empathic compassion of our inter-subjectivity is more than inference concerning another's beliefs and desires; still less is it mere behavior-prediction. It is a voluntary co-experience as if of the other person's mind." (Gunkel & Wales, 2021, p. 479)

8. "A comprehensive definition of empathy recognizes 3 dimensions of empathic engagement: cognitive empathy, or mentalizing, which pertains to the recognition and understanding of the emotional states of others; emotional empathy, or affective sharing, which involves resonating with others' emotional experiences while maintaining self-other differentiation; and motivational empathy, often termed empathic concern or compassion, which encompasses feelings of concern for another's welfare and a readiness to act to enhance their well-being" (Rubin et al., 2024, p. 2)

9. "... created persons exist for relationships. Without relationships, creaturely persons would not cease to exist, but by refusing relationships of self-gift they would live as less than the persons they are." (Gunkel & Wales, 2021, p. 478)

10. "Alles wirkliche Leben ist Begegnung" (Buber & Lang, 2021, p. 33)

11. "We are shaped and motivated by community and by the stories, symbols, values, and practices we share with others, who, in turn, make us who we are." (Segessenmann et al., 2025, p. 187)

12. "Indeed, humans can only express themselves and their experiences in language—whether verbal or bodily—in terms of the words and concepts that they learn in a historical narrative community." (Gastmans et al., 2024, p. 792)

13. "Such virtue—grounded on a relational anthropology of human–technology relations—is the basis of any practi- cal notion of human freedom and morality around which we can organize our liberal, democratic, and plural socie- ties." (Segessenmann et al., 2025, p. 187)

14. "Stellen wir hier an die Spitze des Geistesbegriffes seine besondere Wissensfunktion, eine Art Wissen, die nur er geben kann, dann ist die Grundbestimmung eines geistigen Wesens, wie immer es psychologisch beschaffen sei, seine existentielle Entbundenheit vom Organischen, seine Freiheit, Ablösbarkeit [...] von dem Bann, von dem Druck, von der Abhängigkeit vom "Leben" und allem, was zum Leben gehört – also auch von der eigenen triebhaften "Intelligenz"." (Scheler, 2016, p. 32)

15. "Und die Frage der Philosophie wie im Grunde jede Frage, die der Mensch sich tausendmal im Lauf seines Lebens vorzulegen hat: was soll ich tun, wie soll ich leben, wie komme ich mit dieser Existenz zu Rande –, bedeutet den [...] wesenstypischen Ausdruck der Gebrochenheit oder Exzentrizität, der keine noch so naive, naturnahe, ungebrochene, daseinsfrohe und traditionsgebundene Epoche der Menschheit sich entwinden konnte." (Plessner, 1928, p. 309)

16. "Understanding is a lifelong labor. It is also one carried out not by isolated individuals but by social beings who perform this cultural labor together and share its fruits. The labor of understanding is a sustained, social project, one that we pursue daily as we build, repair and strengthen the ever-shifting bonds of sense that anchor our thoughts to the countless beings, things, times and places that constitute a world. It is this labor that thinking belongs to." (Vallor, 2024b)

17. "Autonomy refers to the capacity to formulate norms, to reflect and to choose which norm to follow, thereby presupposing self-consciousness and freedom. Human agency is autonomous;" (Gastmans et al., 2024, p. 791)

18. "Responsibility is associated with autonomy, freedom and awareness of duties; as such, human beings are responsible. Moral responsibility in any sense cannot be allocated or shifted to 'autonomous' technology." (Gastmans et al., 2024, p. 791)

others (Gastmans et al., 2024, p. 791) [19]; (Segessenmann et al., 2025). Responsibility is not understood as merely fulfilling one's duties, but rather as the ability to respond to the needs of others (Gastmans et al., 2024). Moral action is therefore a combination of thinking, compassion, and commitment to oneself (Gastmans et al., 2024; Segessenmann et al., 2025).

2.6. Embodied *Limited / Vulnerable / Temporal*

Human beings not only have a body, they are their body, their physical form (Plessner, 1928); (Gastmans et al., 2024, p. 790) [20]. This is a unity of body and spirit (Fuchs et al., 2024). It is through the body that human beings encounter the world, and thus it forms the basis of our existence (Crawford, 2015); (Fuchs, 2024, p. 12-13) [21]; (Fuchs et al., 2024, p. 212) [22]. Although the body allows us to be present in the world, it also limits us, making us dependent, vulnerable, and ultimately mortal (Calo, 2024). It is precisely this limitation that is essential to being human (Calo, 2024; Fuchs, 2024; Lawrence, 2024). It reveals a dignity that must be respected (Calo, 2024, p. 223) [23]. It makes it clear that recognition, action, understanding, and every encounter are physically mediated (Crawford, 2015; Fuchs et al., 2024). Only as physical beings can humans be there for others and convey warmth and empathy to their counterparts (Fuchs, 2024).

19. "From developmental psychology, we learn that self-reflection and elationality should be viewed as fundamental human qualities. Humans ask themselves what and how they want to be and then act accordingly. ... humans retain free will and a conscientious capacity that guides them to establish interpersonal relationships" (Gastmans et al., 2024, p. 791)

20. "The embodied and relational nature of human beings implies that they do not have a body; rather, they are their body." (Gastmans et al., 2024, p. 790)

21. "Nur als verkörperte, leibliche Wesen sind wir aber auch füreinander wirklich. Eine Kommunikation oder Empathie zwischen Gehirnen gibt es nicht, auch wenn Neurowissenschaftler das gerne behaupten. Empathie erlernen wir nur im leiblichen Kontakt mit anderen, in der "Zwischenleiblichkeit", wie Merleau-Ponty sie nannte. Und wir verstehen andere nicht erst durch eine "Theorie des Geistes" (Theory of Mind), wie die gegenwärtige Entwicklungspsychologie annimmt, sondern bereits intuitiv anhand ihres leiblichen Ausdrucks, ihrer Gesten und ihres Verhaltens." (Fuchs, 2024, p. 12-13)

22. "A human being always and conjointly is a living body [Leib] [...] and has this living body as this physical thing [Körper]" (Plessner 1970, 34–35, emphases original, insertions from us). In our everyday experience, the lived body is the default: "Our own body," writes physicist-philosopher Michael Polanyi, "is the only thing in the world which we normally never experience as an object, but experienced always in terms of the world to which we are attending from our body. It is by making this intelligent use of our body that we feel it to be our body, and not a thing outside" (Polanyi 1983, 16). The key feature of a lived body is thus its "transparency," as Dorothée Legrand writes: The body is transparent in the sense that one looks through it to the world. At this level, pre-reflective bodily experience is precisely the experience of the world as given through the 'transparent body.' The latter is not perceived as an object but experienced specifically as a subject perceiving and acting, that is, in-the world (Legrand 2007, 504)." (Fuchs et al., 2024, p. 212)

23. "The limits that define our nature, including the ultimate limitation of death, are essential to our humanness. To respect human dignity is to respect these limits. They are not to be overcome but taken as a site of moral reflection about what it means to live well as limited creatures of a certain sort." (Calo, 2024, p. 223)

All of the characteristics described above have one fundamental thing in common. They are not simply given to human beings. In order to preserve them, they must be actively lived and nurtured. Being human is therefore an ongoing process of working on oneself (Comer, 2015; Vallor, 2024a).

3. Examining Interactions (Step 2)

The example of the New York Times (Hill & Valentino-DeVries, 2025) shows that AI systems interact with humans on a personal level. Therefore, based on the six dimensions of the Humatrix identified in **Step 1** (see Section 2), this section examines how the Base AI System interacts with these dimensions. Since the nature of the potential involvement of individual dimensions and their relevance may vary depending on the application context of the Base AI System, a specific use case from the field of psychiatry was selected for **Step 2**. Psychiatric documentation provides a particularly suitable context for examining such interactions, as it constitutes a recurring professional practice in which the support of the Base AI System can overlap with dimensions of the Humatrix.

3.1. Use Case Description

The selected use case focuses on report generation in a psychiatric context, specifically on initial consultations and progress meetings (henceforth referred to as *Psychiatric Session Reporting*). This type of reporting is an integral part of clinical practice in psychiatry and goes beyond purely administrative or legal documentation.

Documentation in Psychiatric Treatment In psychiatric treatment, documenting an initial consultation or a progress meeting goes beyond purely administrative requirements. Unlike many other medical disciplines, psychiatry relies primarily on clinical encounters with patients as its central diagnostic and therapeutic instrument. Patient consultation and its subsequent documentation are the most important means by which symptoms and clinically relevant patterns are identified and interpreted.

In addition to structured elements such as family history, current medication, or living situation, Psychiatric Session Reporting includes the assessment and description of the patient's mental state and psychopathological findings. This requires continued interpretation of verbal and nonverbal expressions and sensitivity to subtle changes in behavior or affect (Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie, 2023; Hausner & Cording, 2017). Such observations cannot be measured directly through technical devices, but must be perceived, interpreted, and integrated in a meaningful way by the physician (Schon, 2011). Therefore, the subsequent writing of the report is an integral part of daily psychiatric practice. Beyond recording clinical information, it allows the physician to recall the encounter. Psychiatric Session Reporting can thus be understood as a practice in which reflective, interpretative, and relational activities are structurally embedded within routine clinical documentation.

Workflow without an AI System

The following steps describe Psychiatric Session Reporting performed without the use of an AI System and serve as an analytical baseline to understand how reflective, interpretative, and relational activities are embedded in the documentation process. A corresponding visualization is illustrated in Figure 3a in Appendix A:

1) The physician conducts the initial consultation or progress meeting with the patient.
2) During the consultation the physician either records the relevant information directly or makes keyword-based notes, some of which may be handwritten.
3) Based on these notes, the physician prepares the report, usually following an organizational template.
4) Recalling the patient consultation in order to incorporate reflective, interpretative, and relational aspects into the report.
5) The physician finalizes and reviews the final report for completeness and accuracy.

Workflow with the Base AI System

The following steps describe Psychiatric Session Reporting with the use of the Base AI System and illustrate how the documentation changes from texts written by the physician alone to system-generated texts. A corresponding visualization is illustrated in Figure 3b in Appendix A:

1) The physician starts the audio recording of the Base AI System and conducts the initial consultation or progress meeting with the patient.
2) During the consultation, the physician may additionally take notes.
3) The Base AI System records and transcribes the consultation.
4) Based on the transcript and any notes, the report is generated by the Base AI System according to the organization's template.
5) The physician reviews the report for completeness and accuracy.

3.2. Interaction Identification

Psychiatric Session Reporting constitutes a recurring clinical practice in which the dimensions of the Humatrix are enacted in concrete ways. This section examines these practices and identifies the points at which the Base AI System interacts with them. Then it analyzes how such interactions may influence the conditions under which the dimensions of the Humatrix are enacted over time.

Lived Expressions of Dimensions of the Humatrix in Psychiatric Session Reporting In Psychiatric Session Reporting, the physician performs situated practices that are not professional skills or techniques in the conventional sense. They cannot be directly learned or trained. Rather, they describe the manner in which the physician relates to the patient and to themselves during the patient consultation and its documentation (henceforth referred to as *Lived Expressions*). These Lived Expressions can therefore be understood as concrete forms in which individual dimensions of the Humatrix can be lived and sustained in everyday clinical practice, particularly in Psychiatric Session Reporting.

Specific examples of such Lived Expressions in Psychiatric Session Reporting:

- Experiencing personal responsibility towards the patient.
- Responding attentively by listening and asking questions.
- Perceiving and interpreting nonverbal signals (gestures, facial expressions, body posture) as well as tone of voice and mood.
- Recognizing inconsistencies between what is said and what is shown.
- Being present in the moment and consciously allowing for pauses or silence.
- Expressing empathy and emotional concern.
- Building and maintaining an atmosphere of trust.
- Drawing on experiential knowledge, perception, and intuition.
- Reflecting on one's own emotional reactions, preconceptions, and uncertainties.
- Integrating individual impressions into a holistic understanding.
- A reserved, responsible approach to interpretations and diagnoses.
- Taking cultural, religious, or linguistic contexts into account.
- Organizing, weighing, and readjusting impressions in the writing process.
- Conscious decisions about what to record and what to leave open.
- Perceiving changes over time.

These examples of Lived Expressions are not intended as a checklist or an exhaustive inventory. Rather, they serve to illustrate where and how dimensions of the Humatrix become visible in the practice of Psychiatric Session Reporting. A possible analytical mapping used for orientation in the subsequent analysis is provided in Table 1 in Appendix B.

Points of System Interaction with the Lived Expressions The process of Psychiatric Session Reporting changes when the Base AI System is used. The documentation process, which was previously carried out mainly by the physician, is partially shifted to the text output generated by the system. The differences in workflows between manual physician documentation and automated system-generated documentation can be seen in Figures 3a and 3b in Appendix A. These changes lead to interactions within the documentation process. Some of these interactions are limited to administrative tasks such as compiling or summarizing information. These interactions primarily affect efficiency and organization and

do not represent a central location where the dimensions of the Humatrix are lived or expressed.

However, interactions also occur in parts of the Lived Expressions that serve as a reflective continuation of the patient consultation. When writing the report manually, the physician recalls the patient encounter. In doing so, they organize their impressions from the consultation, weigh interpretations, and draw on their perceptions and emerging intuition. In addition, they decide which aspects of the conversation should be written down and which should remain open. The writing itself serves as a place of reflection, in which the patient consultation is relived and interpreted. Beyond this, it is not only the process itself that changes, but also the role of the physician, who shifts from being the primary author to being the evaluator or reviser of a pre-formulated text.

Interactions with Lived Expressions do not only occur during the writing process, but can also arise during the patient consultation. The knowledge that the entire conversation is being recorded and will later be processed by the Base AI System can have a subtle effect on attention, tolerance of pauses or silence, and sensitivity to nonverbal signals. Furthermore, awareness of being recorded can influence the conditions under which trust can be established and maintained. Both the physician and the patient can adjust their manner of expression, their openness, or their handling of uncertainties.

These interactions do not have an immediate or isolated influence on the dimensions of the Humatrix. Their relevance becomes clear once the use of the Base AI System has become established in daily and routine documentation.

Habitual Changes in Lived Expressions of the Dimensions of the Humatrix The routine use of the Base AI System in Psychiatric Session Reporting establishes recurring patterns of interaction that can, over time, affect the conditions under which certain Lived Expressions of particular dimensions of the Humatrix are performed. This is not a matter of clinical competence or documentation quality, but rather about the gradual change in the way these Lived Expressions are practiced and maintained in clinical routines. Within the Humatrix, the relational and autonomous dimensions are particularly exposed in Psychiatric Session Reporting, as their realization is structurally linked to those Lived Expressions that take place in the documentation process itself.

Other dimensions of the Humatrix remain relevant, but are not exercised with the same regularity or centrality in documentation practice. Their Lived Expressions are more closely linked to the immediate encounter or to normative contexts and are therefore not explored in depth here.

Relational Dimension of the Humatrix
The relationality in Psychiatric Session Reporting does not end with the patient conversation, but continues in the act of documentation. By recalling the consultation, the patient remains present to the physician as a concrete counterpart even beyond the encounter. If this relational continuity is regularly interrupted by system-generated texts, the way in which it is practiced may change. Critically reading the generated report then focuses more on responding to the predetermined representation of the encounter than on reworking it. Increasing reliance on the system's output can diminish the ability to be present, attentive, and responsive in the moment of the conversation. In addition, sensitivity to nonverbal cues or inconsistencies may be reduced. Since relationality is a fundamental way of relating to one's counterpart and is not purely a professional skill, these changes can also manifest themselves outside the clinical context. For example, in everyday situations, such as how one perceives one's counterpart and reacts to the other.

Autonomous Dimension of the Humatrix
Autonomy in Psychiatric Session Reporting is achieved through the exercise of judgment in organizing impressions, integrating experiential knowledge and intuition, and consciously keeping meanings open. When reports are routinely presented as system-generated texts, the focus shifts from the physician's independent formulation to validation or correction in response to the system's pre-written text. Over time, such changes can influence how decisions are made in general, even outside of clinical practice. Independent judgment, interpretation, and intuition practices may be diminished.

The impact on the dimensions of the Humatrix in Psychiatric Session Reporting does not result from automation by the Base AI System as such, but from its repeated and increasingly habitual interaction with Lived Expressions through which the relational and autonomous dimensions of the Humatrix are realized.

4. Designing Pro-human AI System Interventions (Step 3)

Building on the dimensions of the Humatrix identified in **Step 1** (see Section 2) and their interactions with the Base AI System analyzed in **Step 2** (see Section 3), this section presents specific design interventions for a Pro-human Prototype AI System. The aim is to maintain the Lived Expressions of the relational and autonomous dimensions of the Humatrix and thus preserve both dimensions.

4.1. From Automated Report Generation to Assisted Text Completion

In order to address the analyzed effects on the relational and autonomous dimensions of the Humatrix, the documentation process is redesigned so that the physician continues to be actively involved in the writing process instead of receiving a fully automatically generated report. They remain the author and bearer of responsibility, while the Pro-human Prototype AI System acts as a supporting aid rather than a replacement tool.

A relevant design paradigm can be found in assisted code completion systems used in software development. Tools such

as GitHub Copilot (GitHub, 2024), OpenAI's Codex (openAI, 2026), and JetBrains AI Assistant (JetBrains, 2025) provide context-sensitive suggestions that must be actively accepted, modified, or rejected, while authorship and responsibility remain with the human user. Applied to Psychiatric Session Reporting, this paradigm shifts the system from automated report generation to assisted text completion. The physician initiates and structures the report based on patient consultation, their own notes, and reflective recollection of the conversation, incorporating intuition and clinical judgment. During the writing process, the Pro-human Prototype AI System analyzes the transcript, any existing notes, and the physician's ongoing text input and explicitly suggests only preliminary text segments (ghost text) without generating a complete report.

In this way, the documentation process remains a place of active wording and judgment formation, rather than being limited to a purely retrospective validation of a preconceived representation of the patient consultation.

Workflow with Pro-human AI System Design Interventions

The following steps describe Psychiatric Session Reporting with the use the Pro-human Prototype AI System, designed to reintroduce human authorship and reflective formulation into the documentation workflow. A corresponding visualization is illustrated in Figure 3c in Appendix A:

1) The physician starts the audio recording of the Base AI System and conducts the initial consultation or progress meeting with the patient.
2) During the consultation, the physician may additionally take notes.
3) The Pro-human Prototype AI System records and transcribes the consultation.
4) The physician initiates and structures the report based on the conversation, the transcript, and the notes.
5) The Pro-human Prototype AI System offers context-sensitive text suggestions during writing
6) The physician iteratively accepts, revises, or rejects them.
7) The physician reviews the report, remaining fully responsible for its content and formulation.

4.2. Reducing Habitual Interaction Effects on the Dimensions of the Humatrix

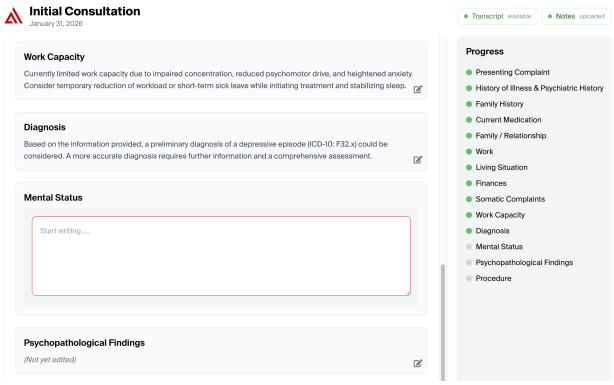
The decisive factor for the pro-human orientation of this design adaptation is not the individual situation, but its habitual repetition in daily clinical practice. Only through this temporal embedding does it become apparent whether and how the proposed intervention stabilizes and maintains the Lived Expressions of the affected dimensions of the Humatrix. By involving the physician in the writing process rather than relieving the physician of this task, assisted text completion supports documentation as a continuation of the interpersonal relationship. Despite the assistance with formulation, the physician is encouraged to actively recall the patient consultation and reflect on it again during the writing process. Since a fully formulated report is no longer generated and the physician acts as the author, the conditions under which attention, presence in the conversation, perception of nonverbal forms of expression, and awareness of possible discrepancies in behavior can be practiced remain intact. At the same time, the system can take over purely administrative text segments, such as family history, current medication, and living situation, as these segments are highly structured and can be reliably extracted from the transcript. This allows the physician to focus more on those Lived Expressions that are central to the relational and autonomous dimensions of the Humatrix during the consultation and, in the best case, even intensify them. The interpretive assessment, the ordering of impressions, the handling of uncertainty, and the decision about what should be recorded or what should be deliberately left open remain entirely with the physician.

Nevertheless, habitual use of assisted text completion can lead to text suggestions being increasingly accepted without reflection, which in turn can influence the practice of the corresponding Lived Expressions. Therefore, the Pro-human Prototype AI System does not completely eliminate all manipulative interactions. Rather, it points to a direction of development that aims to reduce systematic displacement effects through habitual use and enable the continued practice of the Lived Expressions of central dimensions of the Humatrix in Psychiatric Session Reporting.

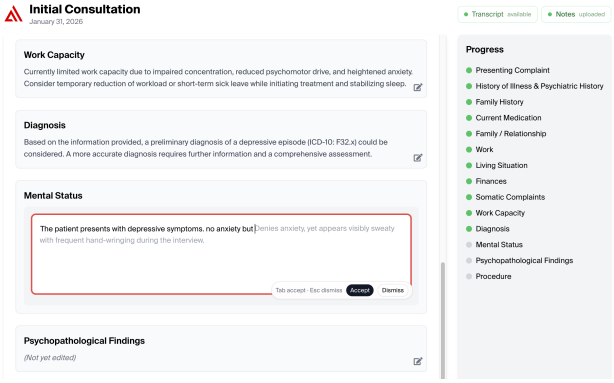
Figure 3c in Appendix B illustrates how the Pro-human Prototype AI System design approach actively reintegrates the physician into the writing process, thus maintaining the conditions under which the Lived Expressions of the relational and autonomous dimensions of the Humatrix can be practiced.

To demonstrate how the proposed pro-human design approach to assisted text completion in Psychiatric Session Reporting should work in practice, a preliminary implementation attempt of a Pro-human Prototype AI System was developed. It shows how the report is no longer generated automatically by the system as a whole, but rather how the physician remains actively involved in the writing process. Figure 2 shows a sequence of the process in which the physician has already completed certain sections of the report, such as presenting complaint, current medication, or living situation and is now moving on to write down the patient's mental status (see Figure 2a). To do this, they enter keywords in the respective field, and the Pro-human Prototype AI System suggests possible text completions based on the transcript and existing notes. The physician can accept, adapt or reject the suggestion through simple interactions (see Figure 2b). In this way, the physician maintains authorship while the system provides support without replacing the whole writing process.

A more detailed description of the preliminary implementation attempt can be found in Appendix C.



(a) The physician has already written some sections of the report using the Pro-human Prototype AI System of assisted text completion, and has now started work on the mental status section.



(b) The physician enters keyword-like assessments into the corresponding input field. The Pro-human Prototype AI System suggests possible text completions based on the transcript and existing notes. These suggestions can be accepted, adapted or rejected by the physician through simple interactions.

Figure 2: Sequence of the Psychiatric Session Reporting process from the preliminary implementation with the pro-human design approach of assisted text completion.

5. Discussion & Conclusion

This thesis examined how AI systems can be designed to be pro-human so that they retain their efficiency gains without diminishing the characteristics that define humans at their core (dimensions of the Humatrix). Using the specific use case of Psychiatric Session Reporting, it was shown that the decisive criterion is not the use of an AI system as such, but rather its routine and recurring use in daily practice. The relational and autonomous dimensions of the Humatrix are particularly exposed in this context. This is evident both in the writing process and in the direct patient conversation, where these dimensions manifest themselves in concrete Lived Expressions such as reflective recall of the patient consultation or responsible and independent judgment. The habitual use of automated report generation changes the conditions under which these Lived Expressions are practiced.

This approach makes the relevant relationships systematically comprehensible in a three-step structure. It combines the identification of constitutive human characteristics (dimensions of the Humatrix) with the analysis of specific interaction points between the AI system (Base AI System) and these dimensions and shows how specific design adjustments can address these interaction points towards a pro-human AI system design (Pro-human Prototype AI System).

The chosen Pro-human Prototype AI System design approach does not represent a complete solution, but rather a design direction that aims to reduce the systematic weakening of the affected dimensions of the Humatrix by maintaining the conditions for the practice of their Lived Expressions. At the same time, risks remain even with the pro-human design direction presented, especially when new patterns evolve through habitual use and text suggestions are increasingly adopted without reflection or further consideration.

Furthermore, the chosen design adaptation of the assisted text completion does not cover all relevant interactions. For example, the potential behavioral changes of the physician and patient that result from transcription during the conversation cannot be resolved. This raises the question whether such interactions can be addressed by design interventions in principle or whether they should be addressed in a different context.

Since this thesis focuses on the specific use case of Psychiatric Session Reporting, it remains deliberately limited. The potential effects of interactions on the affected dimensions of the Humatrix described above are plausible changes resulting from the structure and sequence of the use case. However, these were not measured or empirically evaluated. The preliminary implementation approach for the presented Pro-human Prototype AI System serves solely to illustrate the concrete design and practical handling of the proposed design interventions for assisted text completion.

6. Outlook & Future Work

One possible direction for future work is the targeted further development of the design approach presented here for the Pro-human Prototype AI System. In particular its optimization with regard to remaining risks, functional implementation, and integration into real application contexts. A comparative study could be conducted to examine different interactions and their effects. Users would have the option to choose between fully automated report generation (Base AI System) and assisted text completion (Pro-human Prototype AI System). The aim is to examine the influence of the two different system designs on the conditions under which the Lived Expressions of the relevant dimensions of the Humatrix can be practiced.

By limiting this thesis to the use case of Psychiatric Session Reporting, a further future direction can be derived. The focus here is on developing criteria for the design of AI systems that are considered pro-human. Such a benchmark could enable a comparative evaluation of systems in terms of their

pro-human orientation and their impact on the dimensions of the Humatrix. In this sense, a pro-human benchmark could serve as a guide for future design decisions for AI systems in general.

Remarks on Supporting Tools

Finally, it should be mentioned that AI tools were used in this thesis exclusively to improve the wording and correct the syntax in the text. Specifically, this concerns ChatGPT and DeepL. Furthermore, Copilot was used to understand the existing code structure and for code support.

References

AlpineAI. (2026). *KI-Anwendungen für kritische Sektoren. Swiss Made. Felsenfest. Mit einer menschlichen Note.* Retrieved 2026-01-04, from https://alpineai.swiss/

AnthropicPBC. (2025). *Claude.* Retrieved 2025-11-22, from https://claude.ai

Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie (Ed.). (2023). *Das AMDP-System: Manual zur Dokumentation des psychischen Befundes in Psychiatrie, Psychotherapie und Psychosomatik* (11., vollständig überarbeitete Auflage ed.). Göttingen: Hogrefe. doi: 10.1026/03157-000

Buber, M., & Lang, B. (2021). *Ich und Du* (Lizenzausgabe, [Nachdruck] 2024 ed.) (No. Nr. 14171). Ditzingen: Reclam.

Burwell, J. M. (2025, April). The AI Efficiency Paradox: Reclaiming Quality Patient Care in an Era of Optimization. *Journal of Medical Systems*, *49*(1), 49. Retrieved 2025-09-28, from https://link.springer.com/10.1007/s10916-025-02183-2 doi: 10.1007/s10916-025-02183-2

Calo, Z. R. (2024, July). AI, medicine and Christian ethics. In B. Solaiman & I. G. Cohen (Eds.), *Research Handbook on Health, AI and the Law* (pp. 219–233). Edward Elgar Publishing. Retrieved 2025-09-29, from https://www.elgaronline.com/view/book/9781802205657/ch13.xml doi: 10.4337/9781802205657.ch13

Cassirer, E. A. (1960). *Was Ist der Mensch?: Versuch Einer Philosophie der Menschlichen Kultur*. W. Kohlhammer.

Comer, J. M. (2015). *Garden City: Work, Rest, and the Art of Being Human*. Grand Rapids: Zondervan.

Crawford, M. B. (2015). *The world beyond your head: on becoming an individual in an age of distraction* (First edition ed.). New York: Farrar, Straus and Giroux.

Dehaene, S. (2021). *How we learn: why brains learn better than any machine ... for now*. New York: Penguin Books.

Devgun, A. (2023). Value, Homo Economicus, and Agent-Based Modeling: The Importance of Re-Evaluating the Value Systems of AI. *SSRN Electronic Journal*. Retrieved 2025-09-28, from https://www.ssrn.com/abstract=4348155 doi: 10.2139/ssrn.4348155

Fuchs, T. (2024). *Verteidigung des Menschen: Grundfragen einer verkörperten Anthropologie* (5. Auflage ed.) (No. 2311). Berlin: Suhrkamp.

Fuchs, T., Aszmann, O., & Dürr, O. (2024). Organisms, Prostheses and the Limits of Cyborgization. *Philosophy, Theology and the Sciences*, *11*(2), 208. Retrieved 2025-10-09, from https://www.mohrsiebeck.com/10.1628/ptsc-2024-0016 doi: 10.1628/ptsc-2024-0016

Gastmans, C., Sinibaldi, E., Lerner, R., Yáñez, M., Kovács, L., Palazzani, L., ... Vandemeulebroucke, T. (2024, November). Christian anthropology-based contributions to the ethics of socially assistive robots in care for older adults. *Bioethics*, *38*(9), 787–795. Retrieved 2025-09-29, from https://onlinelibrary.wiley.com/doi/10.1111/bioe.13322 doi: 10.1111/bioe.13322

Gehlen, A., & Rehberg, K.-S. (2016). *Der Mensch: seine Natur und seine Stellung in der Welt* (No. 89). Frankfurt am Main: Vittorio Klostermann.

GitHub. (2024, February). *Copilot.* Retrieved 2025-12-20, from https://github.com/features/copilot

Google. (2025). *Gemini.* Retrieved 2025-11-22, from https://gemini.google.com

Gunkel, D. J., & Wales, J. J. (2021, June). Debate: what is personhood in the age of AI? *Ai & Society*, *36*(2), 473–486. Retrieved 2025-10-03, from https://link.springer.com/10.1007/s00146-020-01129-1 doi: 10.1007/s00146-020-01129-1

Hausner, H., & Cording, C. (2017). Aufklärung und Dokumentation in der Psychiatrie. In H.-J. Möller, G. Laux, & H.-P. Kapfhammer (Eds.), *Psychiatrie, Psychosomatik, Psychotherapie* (pp. 2953–2966). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved 2025-12-21, from http://link.springer.com/10.1007/978-3-662-49295-6_96 doi: 10.1007/978-3-662-49295-6_96

Hill, K., & Valentino-DeVries, J. (2025, November). *What OpenAI Did When ChatGPT Users Lost Touch With Reality.* Retrieved from https://www.nytimes.com/2025/11/23/technology/openai-chatgpt-users-risks.html?unlocked_article_code=1.3U8.3A1u.ZAX9W46WWg-A&smid=url-share

Ihde, D. (Ed.). (2010). *Technology and the lifeworld: from garden to earth*. Bloomington: Indiana University Press.

JetBrains, s. (2025). *JetBrains AI Assistant*. Retrieved 2025-12-20, from https://www.jetbrains.com/ai-assistant/

Lawrence, N. (2024). *The atomic human: understanding ourselves in the age of AI*. London: Allen Lane.

Olson, P. (2025). *Supremacy: AI, ChatGPT and the race that will change the world* (Paperback edition ed.). London: Macmillian Business.

openAI. (2022, November). *Introducing ChatGPT*. Retrieved 2024-06-01, from https://openai.com/index/chatgpt/

openAI. (2026). *Codex*. Retrieved 2026-01-03, from https://openai.com/de-DE/codex/

Plessner, H. (1928). *Die Stufen des Organischen und der Mensch: Einleitung in die philosophische Anthropologie*. Berlin Leipzig: Walter de Gruyter. doi: 10.1515/9783111537429

Rashid, A. B., & Kausik, M. A. K. (2024). AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. *Hybrid Advances*, *7*, 100277. Retrieved from https://www.sciencedirect.com/science/article/pii/S2773207X24001386 doi: https://doi.org/10.1016/j.hybadv.2024.100277

Ross, W. (1956). *Aristotelis De anima*. E Typographeo Clarendoniano. Retrieved from https://books.google.ch/books?id=zCC8s_T-07UC

Rubin, M., Arnon, H., Huppert, J. D., & Perry, A. (2024, June). Considering the Role of Human Empathy in AI-Driven Therapy. *JMIR Mental Health*, *11*, e56529. Retrieved 2025-09-30, from https://mental.jmir.org/2024/1/e56529 doi: 10.2196/56529

Scheler, M. (2016). *Die Stellung des Menschen im Kosmos*. S.l.: Eisenbrauns.

Schon, D. (2011, October). The Reflective Practitioner: How Professionals Think In Action. *The Journal of Continuing Higher Education*, *34*. doi: 10.1080/07377366.1986.10401080

Segessenmann, J., Stadelmann, T., Davison, A., & Dürr, O. (2025, February). Assessing deep learning: a work program for the humanities in the age of artificial intelligence. *AI and Ethics*, *5*(1), 1–32. Retrieved 2025-09-29, from https://link.springer.com/10.1007/s43681-023-00408-z doi: 10.1007/s43681-023-00408-z

Shneiderman, B. (2022). *Human-Centered AI*. Oxford University Press. Retrieved from https://doi.org/10.1093/oso/9780192845290.001.0001 doi: 10.1093/oso/9780192845290.001.0001

Tegmark, M. (2018). *Life 3.0: being human in the age of artificial intelligence* (First Vintage books edition ed.). New York: Vintage Books.

Vallor, S. (2016). *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford University Press. Retrieved from https://doi.org/10.1093/acprof:oso/9780190498511.001.0001 doi: 10.1093/acprof:oso/9780190498511.001.0001

Vallor, S. (2024a). *The AI mirror: how to reclaim our humanity in an age of machine thinking*. New York (N.Y.): Oxford University press.

Vallor, S. (2024b, June). The Thoughts the Civilized Keep. In *The AI Mirror* (1st ed., pp. 102–132). Oxford University PressNew York. Retrieved 2025-10-09, from https://academic.oup.com/book/56292/chapter/445317996 doi: 10.1093/oso/9780197759066.003.0005

Verbeek, P.-P. (2005). *What Things Do: Philosophical Reflections on Technology, Agency, and Design* (R. P. Crease, Trans.). Penn State University Press. Retrieved 2025-10-09, from http://www.jstor.org/stable/10.5325/j.ctv14gp4w7 doi: 10.5325/j.ctv14gp4w7

Appendix A.
Comparison of the different Workflows in Psychiatric Session Reporting

The workflow visualizations provide a structural comparison of Psychiatric Session Reporting across the three documentation processes: without the use of an AI system, using the Base AI System, and using the modified Pro-human Prototype AI System. They show where the differences between these processes lie and how responsibility, authorship, and reflective practices are distributed differently across the reporting process. The visual comparison supports the analytical argument by illustrating changes in the process and their impact on the Lived Expressions of the affected dimensions of the Humatrix.

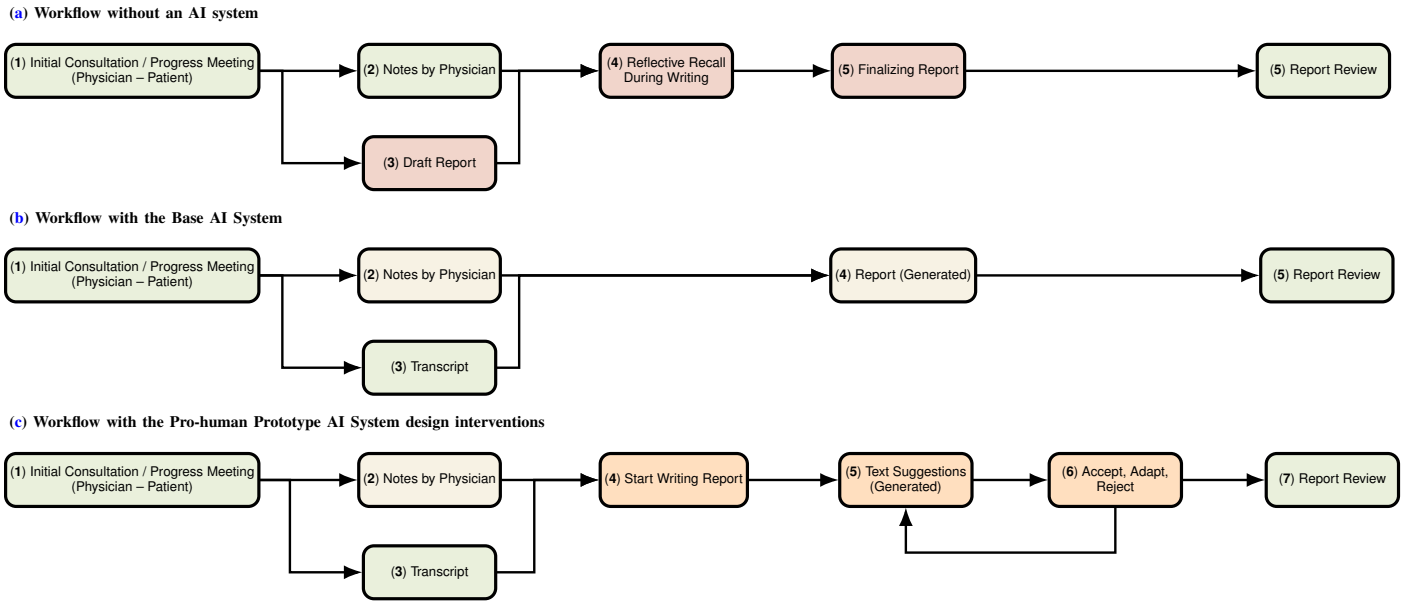


Figure 3: Comparison of Psychiatric Session Reporting across the three documentation processes: (a) workflow without an AI System, (b) workflow with the Base AI System, and (c) workflow with Pro-human Prototype AI System design interventions. Color coding: ▭ common to all workflows, ▭ performed without an AI system, ▭ supported by the Base AI System, ▭ enabled by the Pro-human Prototype AI System.

Appendix B.
Illustrative Mapping of Dimensions of the Humatrix and Their Lived Expressions

The following table provides an illustrative analytical orientation for how selected Lived Expressions observed in Psychiatric Session Reporting can be related to different dimensions of the Humatrix. It serves to clarify the conceptual distinctions used in the analysis rather than to define fixed categories or evaluation criteria.

Dimension of the Humatrix	Exemplary Lived Expressions in Psychiatric Session Reporting
Relational	Experiencing personal responsibility toward the patient; attentive listening and questioning; being present in the moment; allowing pauses or silence; expressing empathy and emotional concern; building and maintaining an atmosphere of trust; perceiving and interpreting nonverbal signals; recognizing inconsistencies between what is said and what is shown; perceiving tone of voice and mood
Social	Taking cultural, religious, or linguistic contexts into account
Transcendent	Reflecting on emotional reactions, preconceptions, and uncertainties
Autonomous	Drawing on experiential knowledge and intuition; tolerating uncertainty; organizing and ordering impressions after the encounter; reflecting on one's own interpretations; integrating impressions into a holistic understanding; perceiving changes over time
Ethically Responsive	Experiencing responsibility toward the patient as moral answerability; reserved and responsible handling of interpretations and diagnoses; conscious decisions about what to record and what to leave open
Embodied	Embodied perception of the patient; temporal awareness of change and continuity

TABLE 1: Illustrative mapping of exemplary Lived Expressions through which dimensions of the Humatrix become visible in the context of Psychiatric Session Reporting. The mapping is not intended as a taxonomy or exhaustive classification. Individual Lived Expressions may relate to multiple dimensions of the Humatrix, reflecting the overlapping and situational nature of their enactment.

Appendix C.
Visualization of the Preliminary Pro-human Prototype AI System Implementation

The following figures (see Figures 4 - 8) illustrate a sequence of the assisted text completion workflow in the Pro-human Prototype AI System. It depicts how documentation is no longer generated as a complete report by the system, but instead emerges through an iterative writing process in which the physician remains the primary author.

The sequence starts with the physician entering brief, keyword-based assessments during report writing. Based on the consultation transcript and any available notes, the system proposes provisional text completions that appear inline as suggestions (ghost text). These suggestions do not replace the physician's formulation, but require explicit interaction. They can be accepted, adapted, or rejected. The illustrated interface represents a preliminary implementation intended to demonstrate the feasibility and interaction structure.

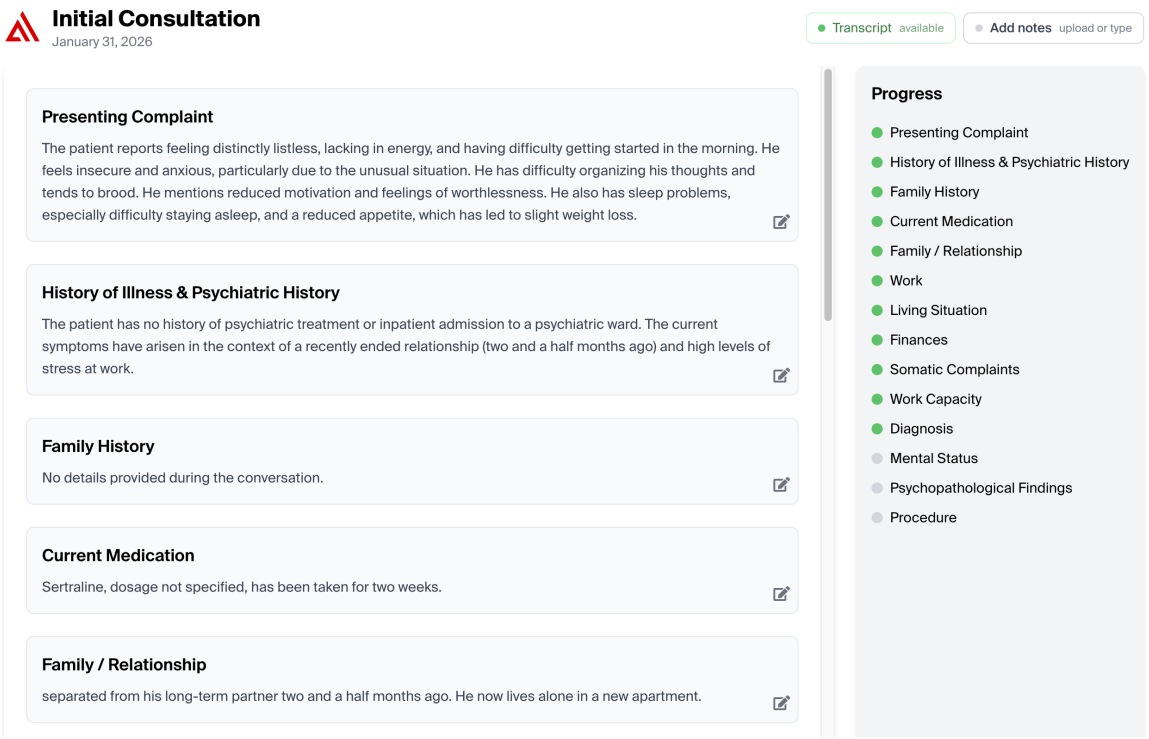


Figure 4: Individual sections of the report have already been written by the physician. However, no supplementary notes have been uploaded to the system, as these sections can easily be obtained from the information in the transcript.

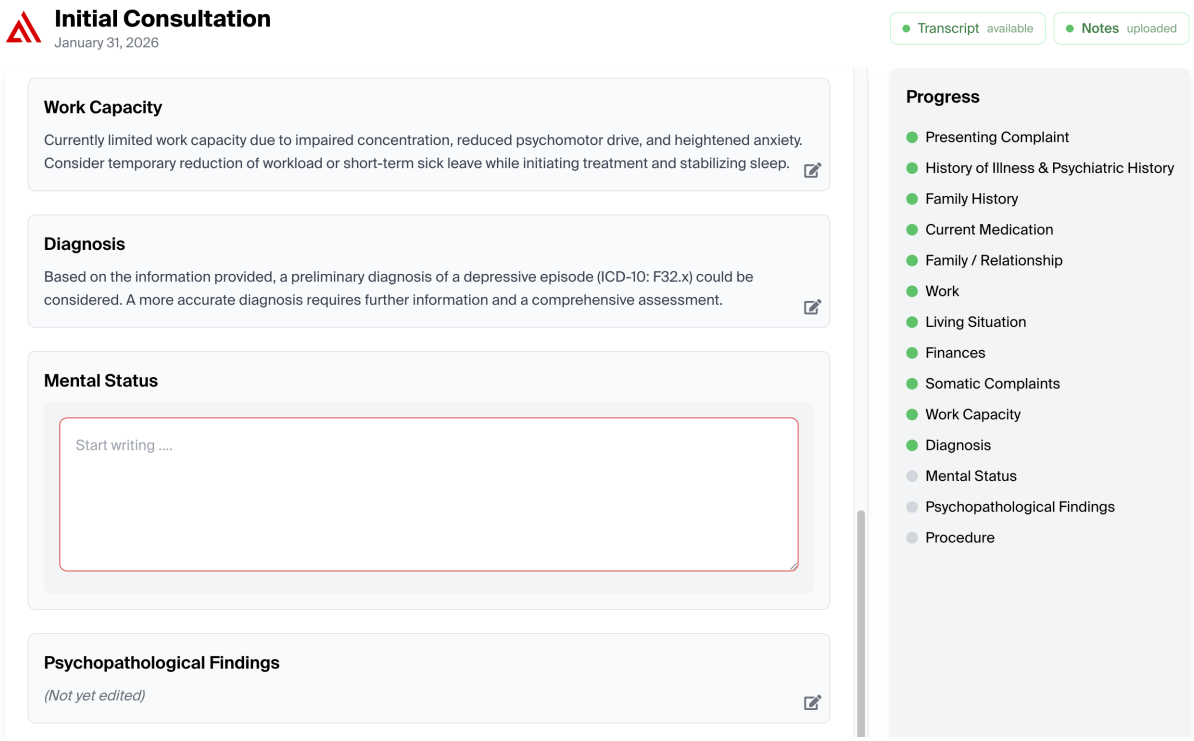


Figure 5: The physician moves on to the section on assessing mental status. In the meantime, additional notes have been uploaded to the system by the physician. Based on their assessments from the patient consultation, the physician starts to write.

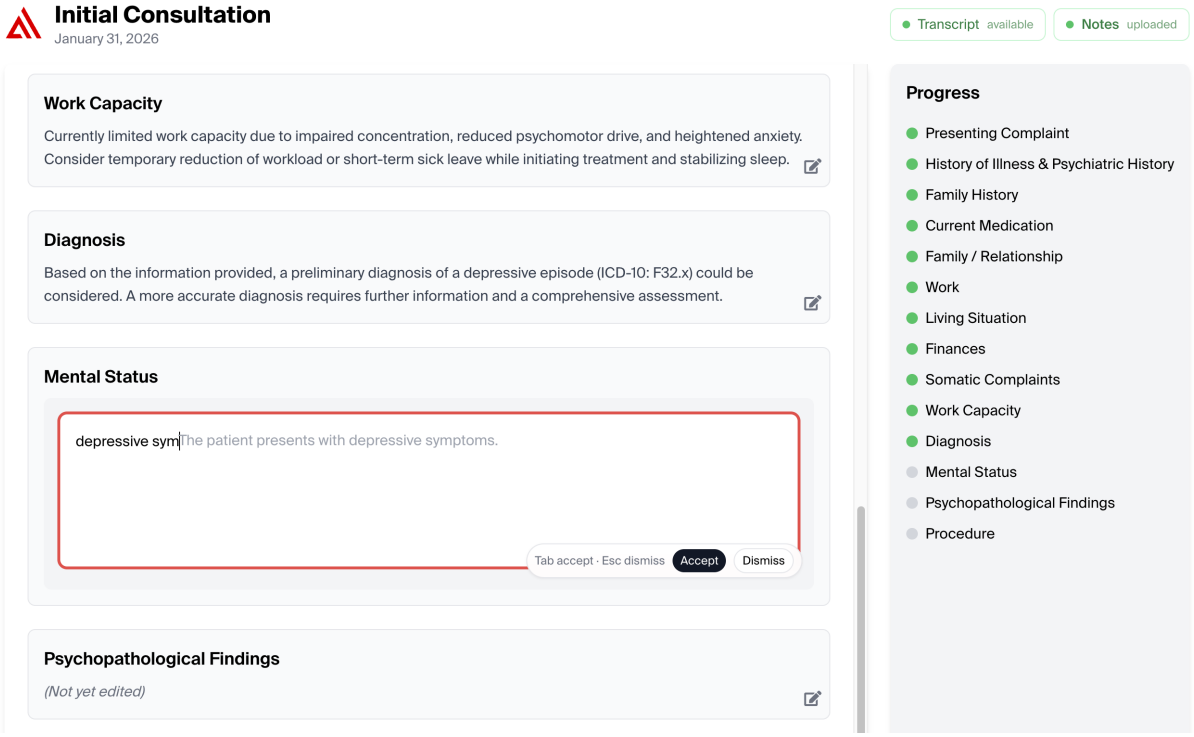


Figure 6: After entering keywords based on their assessment, the physician receives text suggestions from the Pro-human Prototype AI System, which they can either accept, modify, or reject through simple interactions.

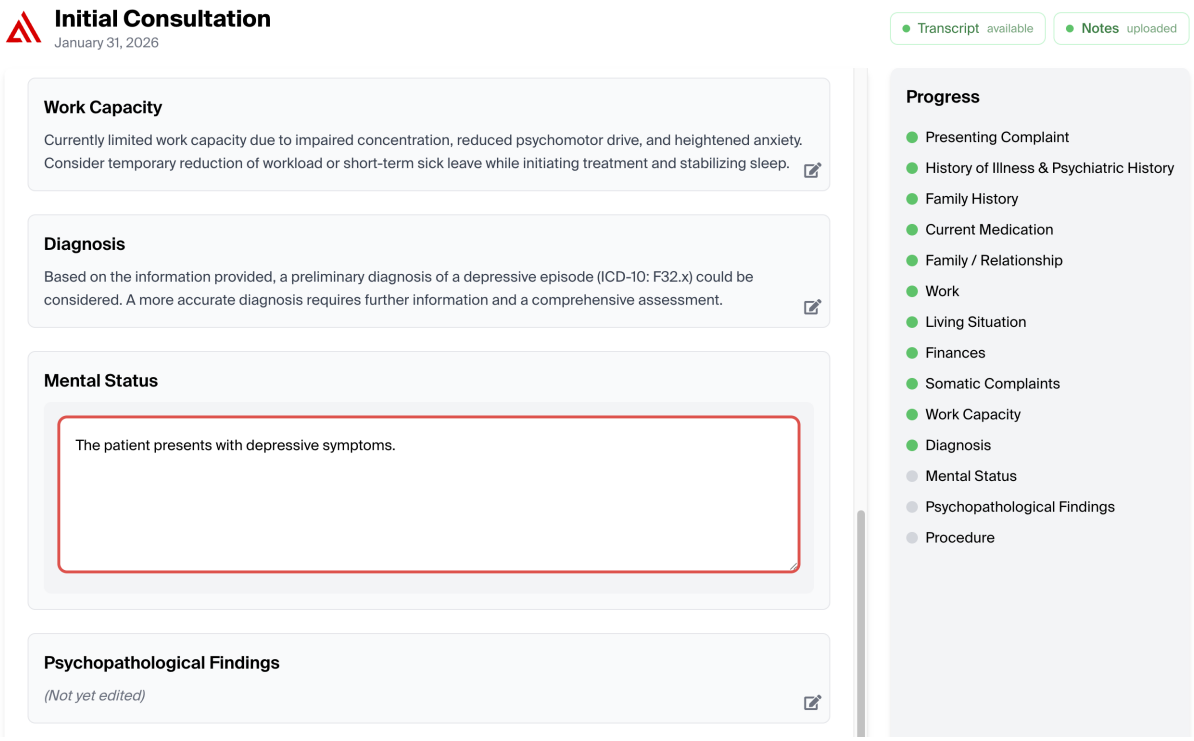


Figure 7: The physician has accepted and used the system's text suggestion for their initial assessment of mental status.

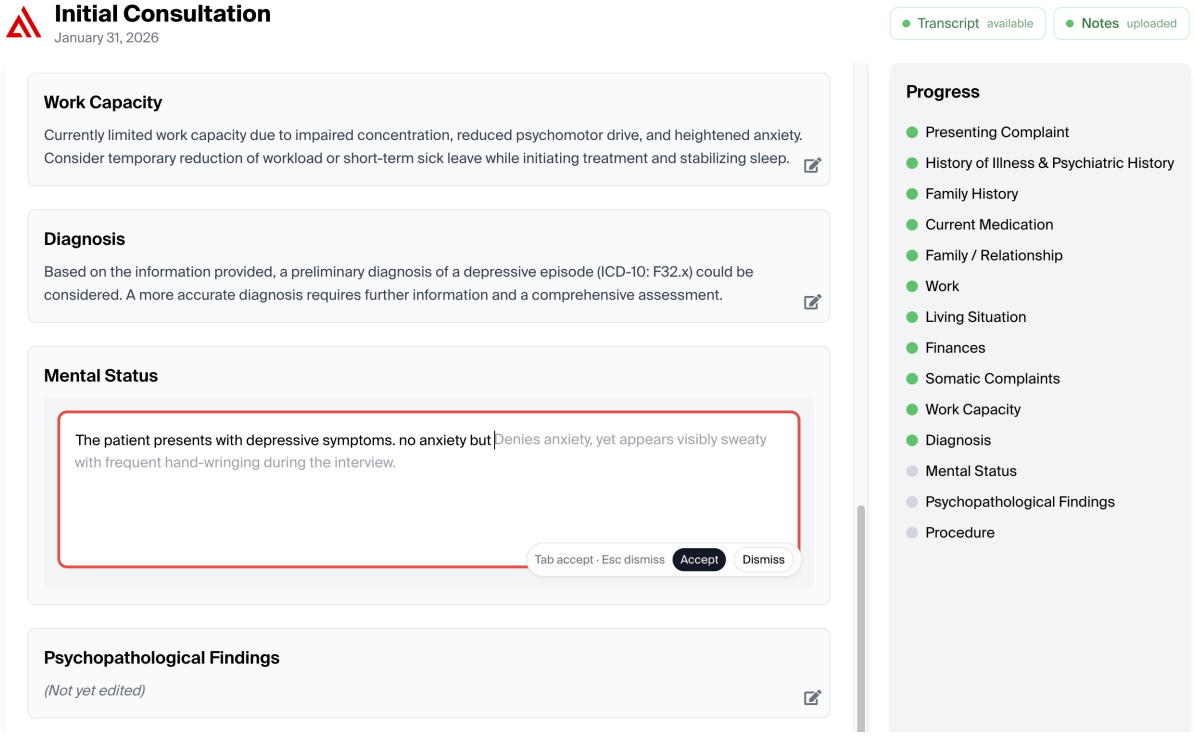


Figure 8: Based on their further assessments, the physician enters the next keywords and again receives context-related text suggestions from the system, which the physician can accept, adapt, or reject. This iterative writing process is continued until the physician considers the section on mental status to be complete, based on the direct patient conversation and its reflective review.