Human-AI collaboration through body posture

Borislav Bonev*
Aalborg University

ABSTRACT

This thesis investigates the design and implementation of a posture-based interactive system called Pose Poet, which maps human body movements to generative poetry. The research explores two primary questions: how mappings between body posture and generative text can foster a sense of ownership and creative agency, and how artificial intelligence can serve as a creative partner that introduces novelty. Drawing on human-centered design, embodied interaction theory, and frameworks for creative agency, the study positions the user as an active collaborator, whose bodily expression informs and shapes the AI-generated poetic output.

Overall, Pose Poet demonstrates how embodied interaction can mediate creative agency in human-AI collaboration. It offers practical insights for designing responsive, intuitive, and ethically conscious co-creative systems. The findings highlight the potential for AI-mediated embodied systems to transform engagement with artistic and educational applications, all through movement.

Index Terms: Embodied interaction—Generative artificial intelligence— Body posture—AI poetry

1 INTRODUCTION

Human creativity has been considered a uniquely human attribute for long time. Expressed though art, literature, and performative art. With the rise of generative artificial intelligence (AI), new opportunities are emerging for humans to express themselves. With the help of machines, humans can co-create faster and easier. That pushes the boundaries of art and technology closer together. one promising domain is the use of bodily movements as an input for creative expression. By using the body as an interface, systems can support embodied interaction and enable users to engage with digital creativity in a new way.

This project and the application Pose Poet explore how human posture can be mapped to generative poetry. It also explores way to facilitate the interaction.

1.1 Problem statements

The problem formulation consists of one main question followed by two research questions. The research questions allow to deep dive into the interaction and help answer the main question.

How can embodied interaction with AI be designed to enable users to co-create generative text, in order to successfully cultivate a strong sense of creative agency in the user?

1.1.1 Research questions

RQ1: How can the mappings between body posture and generative text outputs be designed to foster a clear sense of influence, ownership, and authorship while supporting intuitive and expressive interaction?

RQ2: What design strategies enable AI to act as a creative partner that introduces surprise and novelty, while maintaining user agency and reinforcing the perception of final authorship?

*e-mail: bbonev22@student.aau.dk

1.2 Thesis overview

The thesis consists of a total of eight sections. Section 1 focuses on describing the project and the problem at hand. Section 3 focuses on presenting the literature concepts that relate to the topic. Section 4 discusses the selected methodologies used to research the topic. Section 5 presents how various development activities have been executed and how research methodologies have been applied. Section 6 aims to present the findings of this study. Section 7 reflects on the findings and speculates on additional areas of exploration. Section **??** summarizes the project as a whole. The last section consists of all literature used throughout the study.

2 THEORETICAL BACKGROUND

2.1 Mixed-methods research

Mixed-methods research is defined as ”research in which the investigator collects and analyzes data, integrates findings, and draws influences using both qualitative and quantitative approaches or methods in a single study or program of inquiry” [28, p.15]. Bryman argues that insights gathered through using both research methods can enhance one another and can bring deeper understanding of the topics at hand . [3, p. 628]

There are four techniques that help a researcher combine qualitative and quantitative research - Triangulation, Offset, Completeness, and Process. To summarize, Triangulation focuses on cross-checking results from one research strategy to another. It is often used when a researcher has gathered large amounts of quantitative data and he has later performed an interview to confirm the quantitative findings. Offset is used in situations where the strengths of the qualitative or quantitative research method can offset the weaknesses of the used qualitative or quantitative approach. Completeness argues that more valid results are achieved by including both methods together. Process assumes that quantitative methods show a fixed picture of the questions at hand, while qualitative methods present more the means behind [3, p.635-638]

2.2 Human-centered design

The main objective of the paper focuses on how users collaborate with an artificial intelligence (AI) system and together a poem is created. It is of a great importance that the user feels as part of the collaborative process. Hence, the system (software product) should be designed with the user in mind. Human-centered design (HCD) complements multiple functional and technical requirements for designing a software system. These processes aim to incorporate the user’s perspective into a usable system. In this study, the goal of HCD is to help design a system in which the user feels as a partner of the AI and as part of the creative process, and not as someone who ”feeds the system”.

3 LITERATURE REVIEW

This section of the paper summarizes all works and theories that relate to the study at hand. Each paper and theory has been carefully selected using various search methods. It has been chosen to supply additional meaning to the objective at hand.

3.1 Literature review

he goal of the literature review is to collect in one all relevant analysis on a given topic or areas of one. A literature review aims to

describe, summarize, and clarify various cases of written reports - critical/analytical, theoretical, empirical, etc. [11, p.205]

There are four types of literature reviews - traditional, systematic, meta-analysis, and meta-synthesis. This project relies on a narrative literature review. Its goal is to gather a volume of literature on a given subject area and draw conclusions about the topics in question [6, p. 4]. The purpose of a literature review is to not demonstrate the author's knowledge about a topic but to also allow one to seek new ways of inquiry, identify recommendations for additional research, or gain methodological insights. A literature review serves as a way to introduce the main assumptions to the reader and to justify why certain issues are being explored. Moreover, it can further rationalize the significance of a problem and can help gain a new perspective on it [24, p.2]

Within this project, three main topics have been identified that relate to this project. The structure of this structure follows Harris M. Cooper's [5] definition of a conceptual literature review: "[...] works relating to the same abstract ideas appear together" [5]. The concepts summarized in this literature review are: Embodied interaction (section 3.2) UX within AI generated content (section 3.3) Creative agency (section 3.4)

3.1.1 Search process

Various search methods have been used to select relevant literature. My process was a combination of two search tactics - building blocks of the study and the pearl-growing effect applied across multiple libraries and search tools such as Aalborg University Library (AUB), Primo research assistant, Google Scholar, etc. I started by defining keywords for each concept of the study. The building blocks can be seen in **??**

TABLE

Block	Keywords
Block 1: Embodied interaction	"Embodied interaction", "Body as interface", "Bodily expression", "Gesture input"
Block 2: UX within AI generated content	"Human-AI interaction", "AI generated content",
Block 3: Creative agency	"Creative agency", "User authorship", "Human-AI authorship", "Co-creation", "Expressive control"

This way, I was able to find peer-reviewed articles relevant for the topics at hand. Within the selected studies, I used the pearl-growing effect to explore other studies that might be relevant. That way, I was able to find additional insights and resources that help support the topic of interest.

3.2 Embodied interaction

The role of the body in human-computer interaction has been widely explored in embodied interaction and somaesthetic design. Kristina Höök [12] emphasizes that bodily experience is crucial for meaning-making, it argues that interfaces can become expressive and affective through their ability to respond to physical nuance. This aligns with Menon's [18] work, From Mind to Machine, which demonstrates how gestural input in image generation systems can result in deeper cognitive-emotional engagement. These insights support the assumption that physical movement, and more specifically posture, can function as more than control and can serve as a creative prompt. Lin Lin's [10] AR-based visual poetry system expands on this idea by enabling users to interact with poetic fragments in the physical space. The project demonstrates how spatial movement and visual immersion can enhance poetic meaning, but shortfalls on poem creation. On the contrary, the current research aims to transform the body into a live, generative interface, where posture maps to poetic content. This introduces questions around movement legibility, gesture-to-language mapping, and the design of embodied affordances for creative systems. Finally, Fagerberg et el [8] analyze how gesture shape, effort, and emotional valence can be used in designing affective input. Their framework provides a useful foundation for understanding how physical expressiveness might be interpreted by a system to affect poetic generation, making it a technical and conceptual foundation for mapping posture to poetic elements such as tone or rhythm. To build a clearer picture, Loke Robertson [17] contribute a critical design-oriented lens through their studies of dancers. Their research focuses on bodily knowledge and improvisational movement as rich resources for interaction design, highlighting the experiential and expressive qualities of motion. By highlighting the felt experience of movement in space, they argue that the body can be a site of both knowing and designing. This resonates with the goals of the current project, which seeks to treat posture as a meaning-making input. Where Fagerberg et al. contribute formal structure, Loke Robertson offer embodied epistemologies as a way to think about posture as performance, intention, and authorship, not just data. Their methodology also opens up for including improvisation and reflection in the interaction process, essential elements in supporting creative agency through bodily interfaces. Together, these works suggest that designing for embodied prompting in AI-mediated poetic generation is not just a technical challenge, but a deeply experiential and interpretive one, thus requiring sensitivity to how bodies move, feel, and mean.

3.3 UX within AI generated content

Nowadays users engage more and more with various creative AI systems, therefore providing a strong user experience becomes critical. Shneiderman [26](1983, 2007) introduced the idea of direct manipulation and creativity support tools, arguing that feedback, responsiveness, and expressive control are essential for meaningful creative engagement. In generative systems, this translates to making users feel actively involved in the process with a sense of achievement in mind, rather than just passively consuming the output. Amershi et al. [1] provide a comprehensive framework of Human-AI Interaction Guidelines, which emphasize feedback, transparency, and control. These guidelines are necessary in a posture-based system where input might be ambiguous or unfamiliar. Similarly, Weisz et al. [30] argue that generative AI applications should be designed to preserve user intent, enable refinement, and clearly communicate the AI's role in the outcome. This notion is also supported in Liao et al. [16] where they study how users question AI outputs. They identify common expectations around why a system made a choice and how inputs affect outcomes. In the context of posture-based poem generation, this suggests that even in a highly embodied, expressive interface, users may seek clues about the system's internal logic. Even more so if they feel disconnected from the generated output. Designing for explainability in creative contexts may not require technical transparency, but it suggests interpretable feedback. For example, showing how posture influences mood, tone, or rhythm. This can support iterative exploration, better UX, higher degree of usability, and co-agency, enabling users to better refine their input and feel a greater sense of authorship over the generative process. A key challenge for embodied prompting is that users are often not familiar with using their body as a creative input. This demands well-designed onboarding, real-time feedback, and system transparency to support a positive UX. Kantosalo Riihiaho [14] note that in co-creative systems, user experience depends heavily on whether the system responds in expected and emotionally satisfying ways. Their model for evaluating co-creative processes offers tools to assess how well users understand the system, feel supported, and

remain engaged throughout the creative flow.

Effort Effort concerts the dynamic of the movement. The force, timing, and flow are categories as what gives the motion its character. These movements can be classified with High arousal suggesting quick and fluent movements, and low arousal suggesting sustained and restricted movement.

Valence This dimension tries to base the emotional spectrum in reflecting positive or negative tones. That is associated with pleasure or displeasure. That would be a guide for the interpretation of the movement. Also, can be the setter of the overall emotional tone for the pose shape.

3.4 Creative agency

A focal point of this project is the question of creative agency. More specifically, how users perceive their authorship, influence, and control when working with AI. Bandura's [2] theory of social cognitive agency highlights intentionality and forethought as key markers of agency. In AI systems such as the creative ones, these can be reduced if users feel the AI is doing the work for them. The aim here is to reclaim creative agency through embodiment, allowing users to "speak" through their body rather than text. This is especially important in light of critiques like Sarkar [25], which warns against overstating user involvement in generative systems. Instead, he advocates for designs that foreground the user's expressive role. Similarly, the IKEA effect [19] illustrates that users value outcomes more when they feel they have meaningfully contributed, which this project hopes to amplify through visible, felt physical input. Menon [18] also reinforces the idea that embodied expression enhances user authorship, as gesture-based control promotes a sense of "being inside" the artwork, not just manipulating it. Csikszentmihalyi [7] concept of flow - deep engagement through balance between challenge and skill - serves as an additional lens for evaluating how users emotionally and cognitively connect with the poetic co-creation process. Finally, creative agency is shaped not just by input, but by interaction design. The system must provide clear, intuitive mappings between body and poem, support expressive feedback, and frame the user as a co-creator rather than a bystander. This research aims to contribute to a design framework for embodied co-creation, focusing on users' creative agency and poetic authorship.

4 METHODOLOGY

This section sums up the specific procedures and methods used to identify, process, and analyze the data that relates to the topic at hand. This study uses mixed-methods with both qualitative and quantitative research processes.

The methods introduced in this section of the study are: pre-task questionnaire, an engagement session with the system, a pre-task questionnaire followed up by a semi-structured interview.

4.1 Participants

The ten participants who took part in the study were gathered in multiple ways. Some of the participants were recruited through social media posts on Facebook and Instagram while others decided to take part in the study due to their interest in the project.

4.2 Research design and materials

This part of the section concerns itself with providing a framework for the study and more specifically, the collection of data. This section reflects on the decisions made regarding the research methods chosen for the data collection. The research design represents the structure of the study, guides how the research methods are executed and how data is analyzed [3, p.40]. Additionally, this section provides an overview of all materials used in the study such as: technical equipment, the interactive system, and questions. To gather an understanding of users' prior experience with AI, about their demographics, and their expectations of the upcoming interaction, a survey in the form of a questionnaire is created. The questionnaire aims to understand how the participants view AI, what is their level of interaction with AI, and how often they interact with AI. To gather relevant data on how the participants interact with the system, a system interaction session is conducted. In these sessions, the participants are given a set of tasks which leads them to engage with the AI in a creative collaboration. To understand the outcome of the process - how users feel about the creative process, to understand the degree of creative agency, and how the system performs, another questionnaire is created. To gain a deeper understanding of how the participants perceive this type of AI collaboration, a semi-structured interview is conducted. This type of mixed-methods research design derives from section 2.2 and section 3.2 and more precisely from Process. Bryman states that quantitative data shows a fixed understanding of the topics at hand, while qualitative methods present more the meaning behind. The main focus of the study is to explore which principles make the people who work with the system feel as much in control of the creative process as the AI agent.

4.2.1 Pre-task questionnaire

Collecting valid data and understanding the pool of participants is highly important when conducting research. The most common and primary data collection methods are surveys. Surveys are created from the following sections - sample of participants, a data collection method, a survey tool, and individual questions which are later seen as data, which are then analyzed statistically. Nayak Narayan list three types of surveys: Computer Administered Surveys, Electronic Mail Surveys, and Web Surveys [20, p.32]. Out of all of them, the Web Survey is the most suitable one in relation to this project. The main sections of the survey include: questionnaire design, distribution, and reporting. The different questions and answering options include text boxes, paragraph texts, multiple choice, checkboxes, scale, grid, etc. The main risks of using surveys within a research project include reliability of the opinion expressed by the participants, sampling, response rate, maintenance of confidentiality, and ethical issues ([20, p.33]. The pre-task questionnaire consists of two key parts. The first part focuses on collecting demographic data about the participants. This will allow one to gain more information about who the participants are, what is their background, etc. The second part of the questionnaire is designed to gather information about previous experience with AI. The data derived from this part relies on the self-assessment of the participants. All of the gathered answers will be analyzed using descriptive statistics.

4.2.2 Demographics

The demographic section of the questionnaire consists of five questions. The first two questions concerning age and gender are single choice questions. The question relating to nationality is a free text answer. The question related to current level of education is also a single choice question. The question related to the English level of the participants is based on a ranking scale from 1 to 5.

4.2.3 Previous experience with AI

The second part of the questionnaire is designed to gather insights about previous experience with AI. The questions are focused around the participants' attitude towards using AI and/or collaborating with it. Other questions revolve around whether the participants have any creative hobbies or not. Additionally, the section contains questions about the frequency of usage of AI, whether the participants use AI for creative tasks and how often. That section of the survey helps one determine what is the general attitude of the individuals towards AI and embodied interaction. It also allows one to make assumptions on the relation between, for example, age and AI experience, etc. The title of the questionnaire is named "Self-reporting on AI interaction". The survey also contains a section with guiding notes beforehand.

It is used to encourage the participants to respond honestly and to assure them that this is not an evaluation of their abilities, but simply just of an informative nature. Moreover, the survey is anonymous which means no emails and names will be collected, which should give the participants the freedom to answer truthfully. Some of the questions are based on the 5-point Likert Scale.

4.2.4 AI-collaborative session

The basis of conducting an iterative session derives from the main topic at hand - what design guidelines should be present in order the participants to feel that they collaborate with the AI to generate a creative body of work. The AI-collaborative session can show how the participants worked with the AI agent and the system.

4.2.5 Technical setup

To execute the session successfully, a computer with the system, and a suitable location has been made available to each participant. The system is a web application and the only requirement for using it successfully is for the laptop to have a working web camera.

4.2.6 Materials

There are no particular physical materials needed to execute the session. However, as part of the interaction, each participant would be presented with short information about the study, what is the goal of the interactive session. Each participant will be given a few minutes (2-5 minutes) to get to know the system. Then, each participant is presented with two tasks - one defined by the author of the paper and the other one designed to give creative freedom to the participant.

4.2.7 Post-task questionnaire

The post-task questionnaire aims to measure four key dimensions: embodied interaction, creative agency, perceived output quality, and overall experience. The sections consist of various statements which the participants have to report to on the 5-point Likert scale. Each section aims to explore each topic at hand. The embodied interaction section questions the participants on how well what they described through their posture was translated into meaningful text. Some of the statements which the participants have to report on are: "I felt that my posture directly influenced the poem generated"; "The interaction felt intuitive and responsive"; "I could express emotions effectively through my body" [8] The creative agency section presents the participants with statements such as: "I felt like a co-author of the poems", "The poems felt personal or meaningful to me", "I was able to guide the creative outcome through my body movements", "I felt like a co-author of the poems" [2, 14, 22]. The creative agency section aims to explore whether the participants feel as a collaborator with the AI. The perceived output quality section focuses on measuring whether the participant relates with the generated creative work, and most importantly, whether the participant feels a sense of ownership and expression of the created poem. This sense is measured through statements such as: "The poems surprised me in a good way", "There was a clear connection between my movement and the poems", "The quality of the AI-generated poems was high" [16]. The last section aims to measure the overall experience of the participants who collaborated with the AI and the system presented to them. This has been achieved through the following statements: "I enjoyed interacting with the system", "I would like to use a similar installation again". This collection of statements was chosen to compare insights from the pre-task questionnaire, for example, attitude towards AI collaboration for creative tasks with outcome of the collaboration. Furthermore, that part of the study balances the cognitive emotional and aesthetic responses, and it allows for further insights on the user experience concerning creative systems [1, 27]

4.2.8 Post-task interview

The pre- and post-task questionnaires can present the participants' interest in the topic, their opinion on the AI collaboration through various statements, but it cannot show the full meaning as to why they feel a certain way. To provide further meaning and gain even more insights on the participants' perception on AI collaboration, a semi-structured interview has been included in this study. This section presents an interview guide with an explanation of the methods involved.

Semi-structured interview The semi-structured interview aims to seek the individual perspectives of interviewees, their thoughts, and their points of view on a given topic. Semi-structured interviews tend to be more flexible in formulating and answering questions compared to qualitative interviews. Qualitative interviews focus on the participants's thoughts and what they have to say. Oftentimes the questions asked during an interview might differ from the original interview guides. This is why semi-structured interviews allow the interviewer to follow the structure of the interview guide while allowing him to rephrase the question, asking a new one, or purposefully not asking one. This allows interviews to collect rich and detailed information from the interviewees [3, p.466-468] During some semi-structured interviews, photographs and/or other materials could be included. Including photographs as stimuli during the interview, can be very useful because it might provide additional context or can take the discussion between the interview and the interviewee in a different direction. Sometimes, presenting additional stimuli during the interview, can place the participant in a more familiar setting. It can help them think in another way, recall details that might have been left out if these stimuli were not present [3, p. 476-477]

Interview guide The interview guide is a list of topics which should be addressed during the interview. It can also be a list of issues that should be discussed. It serves as a guide for the interviewer to obtain information from participants while leaving room for flexibility and changes [3, p.469] The questions for the interview should be formulated with the participants and their experience in mind. Some guidelines for writing interview questions are: not formulating the quotations in a misleading manner, they should not be too obvious to answer, the language used should be close to the language level of the participants, the questions should be relevant to the topic at hand. It is important that the interviewer is prepared for questions from the participants - on the nature of the interview, the importance of the topics, etc. Questions should be revised and reformulated when necessary until the interview guide and the interview structure are finalized [3, p.470-471]

Interview in this study The interview in this study will be conducted after the participants have completed the post-task questionnaire.

4.3 Procedure

In this section, the procedure of the data gathering will be described. The whole procedure lasted between 45 minutes to an hour. The procedure was organised according to the following schedule:

1. Pre-experience questionnaire: Captured participant background, creative habits, and expectations. (10 minutes)

2. Short introduction to the system: introducing the system and describing the structure of the collaborative session. (5 minutes)

3. AI-collaborative session: Participants explored the system using free-form body movement, followed by light task constraints to examine how intention influences interaction. (10 minutes)

4. Post-experience questionnaire: Focused on user perceptions of the interaction, authorship, and output. (15 minutes)

5. Semi-structured interviews: Provided deeper insight into affective, creative, and interpretive experiences. (10 - 15 minutes)

Participants experienced the system in a controlled setting and completed both a pre-experience. Participants were invited to explore the system by using their bodies and postures in order to generate poetry. Short instructions were given, so participants can get familiar with the system, and allowing them to engage intuitively and creatively. The location chosen for the session was quiet in order to limit the possible distractions and to support higher levels of engagement and immersion with the system. After completing the interaction with the system, the participants completed a post-experience questionnaire. Afterwards, all participants took part in a short semi-structured interview.

4.4 Data analysis plan

This mixed-methods approach allowed for both quantitative and qualitative insights into the user's interaction with the generative poetry installation, following best practices in HCI and co-creative system evaluation [**?**, 13] (Höök et al., 2003). In short, the quantitative data in this study consists of the data gathered through the pre- and post-experience questionnaire, while the qualitative data consists of elaborating questions in the post-experience interview.

4.4.1 Data from pre-questionnaire

The data collected in the survey 4.4.1: The data above has been

Demographics	**Experience with AI**
Age, gender	Frequency of usage, Likeliness to use AU with creative tasks

Table 1: Pre-survey interest areas assessed in the study.

analyzed using descriptive statistics. The results from the data are presented in Section 6 (Results) of the paper.

4.4.2 Data from the the post-experience questionnaire

The figure below presents all the areas of interest in the post-experience questionnaire and lists of statements within each section. The collected data from the post-experience questionnaire has also

Area	**Intrest**
Embodied interaction	How well participants felt their body movement influenced the generated output
Creative agency	Perception of authorship
Perceived output quality	Perceived satisfaction with the output
Overall experience	general usability, engagement, and enjoyment

Table 2: Post-survey interest areas assessed in the study.

been analyzed using descriptive statistics. In the Results section 6 of the paper, there is a comparison of participants' answers between the "Experience with AI" statements and "The Overall experience" statements.

Area
Why was the experience good or bad...
What did you enjoy the most about the experience..
How clear was the connection between your body movement and output...
In what other settings you think the system will be meaningful...

Table 3: Post-survey interest areas assessed in the study.

4.4.3 Data from interview

The data gathered through the semi-structured interview was only qualitative: The analysis performed for the qualitative data was done through Content analysis. The key finding can also be found in Section 6.

5 Application design

This section describes how the application works. It also concerns itself with the mapping of the movement and it's translation to a prompt. It also focuses at the technology choices that were made in order to build the application for this experiment. This section provides an overview of how the application functions. It covers the process of mapping user movement and its translation into prompts that are used for poetic generation. The section also looks into the technology choices made to implement the system. The discussion highlights both the conceptual workflow and the practical considerations involved in building an interactive application for this experiment.

5.1 Overview of Human pose estimation

Human pose estimation is a computer vision technique used to reconstruct the positions of human body parts from visual input such as a photo or a video. That is achieved by identifying anatomical landmark points of the human body, also called keypoints. These points map that to specific joints on the human body such as elbows, knees, hips and shoulders. That allows to later build a "skeleton" structure, which is a simplified representation of the human body with the help of the keypoints [31]. Although this representation is highly generalized, it is still effective in telling the body position. This representation is important as it can be the first step to translate the physical movement and position to a digital interpretation that can be further used. In their survey, identify two main methodological approaches to human pose estimation as top-down and bottom-up. They also outline the dimensional categories: 2D and 3D estimations. Additionally, the field distinguishes between single- and multi-person scenarios. In the top-down approach, first is employed a person detection that identifies the boxes for each individual in an image. After that each single person pose is estimated within that box. That provides better results as isolating each person makes the process simpler to identify a single set of keypoints. That also can prove to be more complex and expensive with the number of people. In contrast, the bottom-up approach, first focuses on identifying the key points and later groups them to make the digital skeleton. That makes it faster and less computationally expensive, however the grouping of the key points to each individual, in the case of a multiple-person case correctly can prove difficult and faulty.

5.1.1 Technology

When utilizing a human pose estimation (HPE) with a webcam, the selection of the models and technology is important in order to achieve a good UX, but also good usability. The integration of human pose estimation (HPE) with real-time webcam input needs a carefully selected model and framework that would support the application in the background. That means ensuring technical feasibility and

optimal user experience. Model selection must balance a number of factors [31]. Some of them include, computational efficacy, latency and ease of development, and maintaining usability.

In their comparative analysis Kalkrani [15] investigated state-of-the-art HPE solutions. They evaluate their performance, benchmarking, adaptability and requirements with the hardware they run on. Some of their findings also highlight that the choice of technology should be also influenced not only by raw accuracy but also the context of application. That means that in the context of this application the real-time human-computer interaction is important. The model should be responsive and should support a smooth interaction, where the user can immerse and recognize themselves in the digital representation. That would also allow the user to feel in control, as they are ensured that the system successfully is able to map their body.

For the present project, MediaPipe coupled with the BlazePose model stands out as a solution that satisfies the requirements. Meaning that high-quality, full body landmark detection, real-time and on devices with processing capabilities that do not require powerful hardware. Looking beyond the efficacy and accuracy that this solution is reported to have, on-device processing also aligns the best with ethical and privacy considerations. By utilising the data locally, the system minimizes risks of sensitive imagery data being transmitted to external third party processing or being stored. In [31] Deep learning-based human pose estimation: a survey, privacy guards in the foundational design of such solutions are extremely important to the user but also align with data and privacy protection legislations around the world.

In summary, the adoption of MediaPipe with the successor of BlazePose in this project addresses the balance of the need for this project but also the responsiveness, usability and ethical responsibility. This technology is therefore well-positioned to support the goals of the system and experiment.

5.1.2 The semantic gap

One of the biggest challenges of this project's development is bridging the semantic gap between the raw numerical representation of the pose data and the higher level abstractions needed for generative AI to create poetry. The kind of representation that poetry requires exists on a more conceptual, symbolic plane. The raw output of the HPE is a set of coordinates for each of the keypoints on the body. From this data, derived features can be determined such as joint angles or calculating the limb velocities. While these abstractions already provide a lot more meaningful information than the raw coordinates, it remains limited to semantic interpretation in a way that would translate to natural poetry. This problem is an example of what researchers call the semantic gap. The distance between low-level data and a higher-level concept, especially when connecting visual information to language in computer vision [9]. As poetry operates on a higher level of meaning, based in narrative, emotion and ideas. As Ollila Jantas [23] define it, poetry is "language... intended to tell a story or express emotion, ideas or states of being". The gulf between quantifiable data that may hint at a story and the narrative, symbolic, themed nature of poetry is wide. A direct, one-to-one mapping from numeric coordinates to poetic lines is not possible. To reach the level of narrative and emotion that poetry requires, the data must first be transformed into more meaningful conceptual representation.

5.1.3 Affective feature extraction and interpretation

As established poetry, by nature, goes beyond factual reporting, but it focuses on metaphors, emotional depth, and abstract concepts. This is why a literal translation as the left arm is bent in a 90 degree angle is sufficient. Although factually accurate, it still binds to literal translation of the data. Affective qualities of the position and movement, could produce nuance that could allow for more interpretations. The underlying emotions and expressive meaning can be translated to a very direct, yet abstract way. The framework proposed in Designing gestures for affective input: [8] offers a productive way forward. Rather than treating gestures as static positions, this model considers how movement properties can encode affective meaning. The framework focuses on shape, effort, and valence. These dimensions become very important because they provide a structured way to interpret the expressed content of the human movement. That is done without many physical description in a try to capture emotion and conceptual nuance.

Shape This dimension describes the changing forms the body makes in space. It can be used to communicate openness, contraction or retreat. This dimension consists of the following paired classifications **Spreading vs Enclosing** This corresponds to the horizontal plane of movement in the shape dimension. Being Enclosed - suggestion of contraction, and spreading - suggestion for openness **Rising vs Lowered** This pair looks into the vertical plane of movement. Rising suggesting uplift, lowered - suggesting grounding.

This matrix would result in four categories for each frame that is used:

1. Spreading Rising
2. Spreading Lowered
3. Enclosed Rising

These categories would allow for a symbolic spatial representation of the position of the body, avoiding the need for prompting with raw coordinates.

5.2 Human-centered design in practice

5.3 AI prompt construction

5.3.1 Model

Since the project is designed as an interactive and real-time application, the model responsible for generating poems must address some criteria. It should be responsive and low-latency. This is important as it can enhance the user experience, but also fosters a sense of immersion. Among today's most widely used models as ChatGPT, Claude, and Gemini-the underlying architectures are well-suited for creative text generation in interactive systems. For this task, the chosen model is Gemini-2.5-Flash. It provides the low latency and responsiveness required for real-time poetic interaction. The capacity of large language models (LLMs) to engage in abstract, metaphorical, and constrained generation is well-documented [29]. This supports the claim that the model can effectively interpret the abstract "affective analyst" input and translate it into poetry. The resulting outputs are not limited to literal descriptions of movement or analysis. This can produce meaningful, emergent poetic expressions that the user can recognize. Furthermore, Gemini-2.5-Flash is a relatively lightweight and resource-efficient model, making it cost-effective. This will be great in scenarios involving large-scale or repeated usage. While more computationally intensive models might achieve deeper or more sophisticated poetic outputs, Gemini-2.5-Flash represents a balanced choice. Meaning that it is powerful enough to generate creative and meaningful poetry, but also efficient enough to support real-time, interactive experimentation. This is why this model positions itself in a "sweet spot" for a project of this size and can serve as a foundation for future, more advanced explorations.

5.3.2 Prompt

Every request to a large language model (LLM) consists of two main components: system instructions and a user-generated prompt. As discussed in Section 5.3.1, the model is expected to interpret input beyond its literal meaning. The system instructions provide

the structural framework for the LLM, while the content of these instructions would vary depending on whether the task is to generate a new poem or refine an existing one. Despite these variations, both forms of instructions follow a common structure with examples from the creation prompt.

In their work Prompt Insights, the authors outline a set of principles designed to enable more effective interactions with LLMs [4]. A representative example prompt begins with the instruction:

"You are an avant-garde poet that directly leverages Principle 16: Assign a role to the large language model."

This role assignment allows the LLM to adopt a specific persona, guiding its interpretive and generative behavior. The prompt then incorporates both positive and negative constraints to shape the nature of the output. For instance:

"Do not describe the pose literally; instead, capture its essence, emotion, or implied story. Be creative and methodical."

Further directives specify the required form:

"Your task is to write a short, evocative, and abstract poem of 4–6 lines." "Your response should only be the poem itself."

These explicit requirements align with Principle 25: Clearly state the requirements [4], ensuring conciseness, clarity, and task alignment. By constraining the output length and format, the prompt reduces ambiguity and enhances reliability.

Another section of the prompt provides contextual grounds:

"The poem should be inspired by a detailed analysis of human movement, including its primary character, shape distribution, and transitions. Do not describe the pose literally; instead, capture its essence, emotion, or implied story. Be creative and metaphorical."

This level of specificity introduces contextual relevance and balances negative constraints with strong positive directives. Such structure enables the LLM to generate outputs that are both imaginative and appropriately scoped.

By leveraging established principles from Prompt Insights [4], users ensure that the LLM can comprehend and effectively execute the task, producing outputs that require minimal post-processing. This not only streamlines the workflow but also facilitates a more responsive system and helps facilitate direct co-creation.

5.4 Application workflow

The application workflow describes how information and interaction flows between the user, the system, and outlines each step of the process. It also depicts how all elements are put together to achieve the poem creation. The workflow consists of:

User interaction The user begins by interacting with the system. The performed movements, the application will capture and interpret.

Movement capture After the initial interaction, the system records five frames of the user's movement. These frames are analyzed and provide the raw data about posture, as described in section 5.1.3

Movement Analysis The captured movement is processed according to the posture analysis methodology described in Section 5.1.1 That provides the translation of the motion into an effective analysis. That includes:

1. Arousal: Providing insights into the intensity of the movement.
2. Transition: Comparing the shape of the postures.
3. Final Shape Valence: Using the last frame to determine the overall emotional valence.

Prompt Assembly The affective analysis, together with the type of prompt is converted into the final prompt for that round. That forms the final text that is required to Gemini.

Poem Generation The assembled prompt is sent to Gemini-2.5-Flash. In return the output is a generated poem that reflects the interpretation.

Output Display The generated poem is displayed to the user as it arrives. That marks the completion of the initial interaction cycle.

Feedback Iteration The user may choose to perform another movement sequence that will help refine the poem. This triggers the same workflow (steps 2–5), but the prompt indicates that the AI should improve or adjust the existing poem instead of generating a new one from scratch. This allows iterative, responsive interaction and encourages expressive exploration.

That forms the basis of the experience. Capturing a few frames allows the user to explore the movement. The structured prompt approach that combines system instruction but also the user-generated affective state allows the generative AI to produce rich, metaphorical, and contextually grounded poetic results.

5.5 UI Pilot Study

Purpose In order to validate the user flow and refine the UI, a small pilot study was conducted. The aim was to confirm the current design or identify areas for improvement.

Methodology **Participants** For this small study, only 2 participants were recruited through word-of-mouth.

Procedure The think-aloud protocol was used in order to collect insights as the users were testing the application.

Key Findings and Iteration A key finding from the pilot study was that participants experienced confusion regarding the nature of the changes in the AI-generated poems after a refinement. While they could notice significant shifts in tone and specific words, more nuanced changes often evaded their attention. This made it difficult for them to understand the direct effect of their input.

To address this issue and improve the sense of control and creative agency, a design interaction was implemented. The application's interface was updated to highlight the new suggestions or outputs, while preserving the old content from the previous interaction. This approach allows users to clearly see the direct result of their refinement, creating a more transparent and intuitive feedback loop that supports iteration.

5.6 User Experience and Interface Design of Pose Poets

The user experience (UX) and user interface (UI) of *Pose Poets* are designed around the central idea of an intuitive and co-creative process between people and AI. The interface acts as an invisible bridge, translating physical expression into AI-generated artistic collaboration. For this reason, the design emphasizes trust, transparency, and experimentation. The approach is grounded in fundamental principles of Human–Computer Interaction (HCI).

5.6.1 Visibility of System Status

A core principle of HCI is that users should always be able to determine the state of the system and understand what it is doing. In *Pose Poets*, this is achieved through the following features:

- **Live pose skeleton overlay:** A real-time digital skeleton is overlaid on the user's video feed, showing the system's perception.
- **Dynamic status text:** A message beneath the main call-to-action (CTA) continuously updates the user on the system's activity.

- **Capture progress bar:** A sequence of dots fills as frames are collected, providing both a status overview and a tangible sense of progress.

Together, these features provide immediate and continuous feedback, addressing Nielsen's first heuristic, *Visibility of system status* [21]. Users do not have to guess as they can see how the system is responding. This kind of transparency is especially critical in AI-driven experiences. The design also reflects Amershi et al.'s guideline to *show contextually relevant information* [1], ensuring feedback is meaningful and well-timed.

5.6.2 Supporting Mental Models

One of the main challenges in HCI for AI systems is helping users form a mental model of how the system operates. Without this, interactions risk feeling like a "black box." *Pose Poets* addresses this challenge through:

- **Difference view for refinement:** When a poem is refined, edits are highlighted with green for additions and red for deletions.
- **Review creation slideshow:** Users can replay the 15 captured pose frames alongside the generated poem.
- **AI prompt and analysis pane:** The detailed prompt generated from the user's movements is made visible, revealing how the AI interprets input.

These features address the user's implicit question of *Why did my AI partner do that?* The difference view shows outcomes of refinements, the slideshow illustrates the system's "story," and the prompt pane exposes the reasoning behind the output. This directly supports Amershi et al.'s guideline to *make clear why the system did what it did* [1].

5.6.3 Direct and Intuitive Interaction

The interface emphasizes control and directness in interaction:

- **Pose as input:** The user's body is the primary input, captured directly by the camera.
- **Contextual CTA labels:** The main button dynamically updates its label to reflect the stage of the creative process.

Using the body as input leverages the HCI concept of *direct manipulation*, where interactions closely map to real-world actions. Dynamic CTA labels provide *feedforward*, helping users anticipate the system's next step.

5.6.4 Aesthetic and Minimalist Design

Finally, *Pose Poets* draws inspiration from contemporary AI interfaces:

- **Minimalistic, dark-themed UI:** A simple pane layout with high-contrast elements.
- **Single prominent CTA:** Keeps focus on the main interaction.
- **Micro-interactions:** Subtle animations, such as the poem "reveal," add polish and delight without distraction.

These choices reflect Nielsen's heuristic of *Aesthetic and minimalist design* [21], ensuring that only task-relevant elements are presented. By removing visual clutter, the UI helps users focus on creation while aligning with interaction patterns familiar from AI-driven products such as ChatGPT or Gemini.

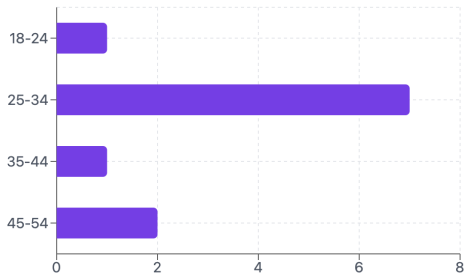

Figure 1: Age distribution

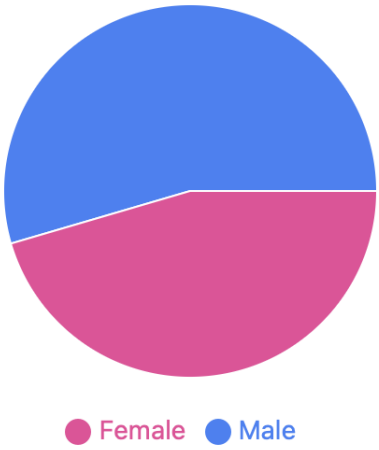

Figure 2: Gender distribution

5.6.5 Summary

By grounding its design in HCI principles and guidelines, Pose Poets enables users to create, manipulate, and explore in collaboration with AI. The interface not only facilitates co-creation but also embodies it, making the collaborative nature of the experience visible, transparent, and intuitive.

6 Results

This section provide an overview of the collected data. It will also include the analyzed data that have been collected and relevant information that can help answer the research questions.

6.1 Data sets

The finalized data set was gathered in an csv file with name "SurveyResponse" , in total there are 20 columns and 12 rows:

Each of the participants is represented in a column.

The answers were mapped where "Strongly agree" was 5 and "Strongly disagree" was 1. The exception was the negatively asked question that follows the opposite scale.

6.1.1 Participants

A total of 11 participants took part in the study. That includes 6 men and 5 women. Visualized at figure 2. The age distribution was diverse with 1 participant was between 18-24 years, 7 participants were between 25-43, 1 participant was 35-44, and 2 participants were between 45-55. The age distribution can be seen at figure 1 The overall engagement with creative activities varied with 4 of

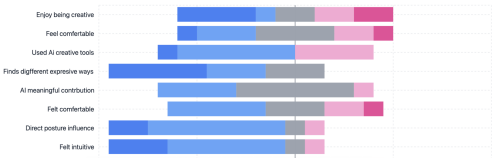

Figure 3: Pre- survey score distribution

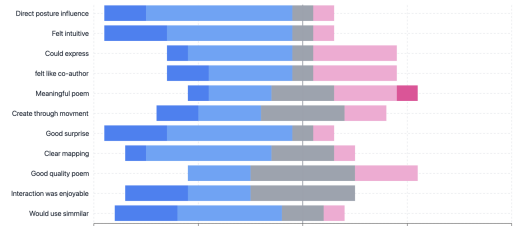

Figure 4: Post- survey score distribution

the participants reported that they engage in creative practices ”frequently” or ”very frequently”. Other 3 of the respondents marked ”never” and the remaining participants ranged between ”occasionally” and ”rarely”. This provided a mix of perspectives from both creative and less creatively involved participants.

6.1.2 Pre- questioner

The pre-experience responses revealed generally positive attitudes towards creativity and experimentation. Most of the participants reported that they enjoy some form of creative tasks and general openness to new modes of self-expression. Comfort with bodily interaction as means of inputs gave mixed responses. Some participants strongly agreed they were comfortable using their body to interact with digital systems, others disagreed, as seen on figure 3. This highlight difference in embodied literacy [12]

Expectations regarding the AI counterpart are barely on the optimistic side. Several participants seemed to agree or strongly agree that Ai could be a meaningful contributor to the creative processes. On the other hand, there were those who were more skeptical or neutral towards that. The same was observed with the attitudes towards poetry. Some of the participants expressed strong comfort, while others were on the disagreeing side. These results suggest that participants that have taken part of the survey have varied attitudes and exceptions.

6.1.3 Post-questionnaire

During the experiment, the participants used their bodies and body postures to generate poems. The post experience responses reveal generally positive engagement with the application, however some variation can be observed. The distribution of the scores can be seen also on figure 4.

Embodied interaction Most participants agreed or strongly agreed that their posture influenced the generated poems, and they report that the interaction behind it felt intuitive. Some of the participants even reported that they were able to express emotions effectively through movement. However, a small group of participants disagreed or remained neutral, suggesting that for some of them, the mapping between body and text was not completely clear.

Creative agency According to the reports, many participants felt like co-authors of the poems and reported that the poems were meaningful or personal. However, there were several participants who gave a neutral or negative response, indicating that there was a difference in the general attitude towards agency between the participants and the AI.

Perceived output quality Most of the participants reported being positively surprised by the poems. Additionally to that, some of them reported that there was a clear connection between movement and text. However, perception of the overall quality varied. Some of the participants found the poems to be less satisfying in their expectations.

6.1.4 Semi-structured interviews

Two participants (P1 and P2) completed short semi-structured interviews following the interaction as described in Section 4. Their responses were coded thematically to complement the findings and data from the questionnaires. Three major themes emerged: authorship and co-creation, clarity of mapping, and the role of surprise and control.

Authorship and co-creation In what they shared, both participants reported some sense of co-authorship. However, it differed in the way they saw that. P1 felt the poems emerged from their own creative intention, with the AI ”Filling the blanks”, supporting what was already their idea. On the contrary, P2 described their contribution more as enabling the system (”it needs me to activate it”). P2 also questioned if whether the output could meaningfully represent personal emotions. These findings support and align with the data in the questionnaires. The participants felt like co-authors, but reveal some underlying skepticism about how deep their contribution extends.

Clarity of mapping The participants reported mixed perception of how their body movements influenced the poems. P1 estimated that the connection was only ”65-75% clear”. That indicates that there was observed responsiveness, but could also point towards their own lack of freedom for bodily expression. P2 acknowledge observed connection, noting that fast movements produced ”wind” in the poem. P2 also noted that these internal states cannot be fully translated though gestures or body position. These findings align survey data, supporting the idea that people find the interaction intuitive but remain unsure about the translated motion.

Surprise and control Surprise was seen differently. P1 reported no meaningful surprises, while p2 found the AI’s phrasing novel and beyond typical human capacity. That points that they saw the system as generator of unexpected ideas. At the same time, both participants expressed concerns about the pacing. P1 described the system as ”fast”, making it difficult to fully engage. P2 even critiqued the limited time frame for expression. These findings reinforce the insights from the questionnaires and observed trends in variability in perceived control. Additionally, highlighting the need for adjustable pacing and transparency in mapping.

Broader user reflections When asked about potential application of such systems, both participants note some practical use cases. P1 proposed medical and rehabilitation context as detection and tracking of recovery after stroke. P2 had a vision for application in education, especially in the settings of non-verbal communication. Both participants could see the benefits of a similar system in used in their respective fields of work.

6.1.5 Summery

Overall participants who were more comfortable with bodily interactions prior the session tended to report higher influence and expressiveness in their post-session result. Authorship is variable with some participants feeling strongly like co-authors, others feeling like a step back.

7 Discussion

7.0.1 RQ1

The results suggest that a clear and intuitive mapping between body posture and generative text is central to the participants’ sense of

ownership. In the post-questionnaire most participants agree that their posture influenced the poems. However, that experience seems to be varied and clear and intuitive mapping might not always be achieved. When the mapping was perceived by the user as direct and responsive,the participants generally reported a strong sense of ownership over the poem. That strongly suggests that the embodied interaction is strongly felt. The pattern can be also observed when the connection was less transparent or visible to the participants. In that case they seem to be more detached from the final output. For example, P2 in the interviews, emphasized skepticism, highlighting that poetry traditionally reflects inner emotions and states. According to P2, these states are not easy to convey through movement.

that aligns prior findings in somaesthetic design [12] [8] which show that bodily expressivity enhances user engagement when the system's feedback is perceived as congruent.

Traditionally, when evaluating a creative agency the user is the creative actor. Nowadays, with the introduction of AI and human-AI creative systems like Pose poet, the actors become two. The concept of creative agency shifts from singular human authorship to a more distributed model.

The user is no longer writing the words needed to form a poem, the artist or user now defines with their input the frame for the poem. The output is also refined by the user like an editor but still a creator. Meanwhile the AI follows the guidance but it also introduces new ideas and surprise, maybe ambiguity. That becomes a catalyst for inspiration and creation of ideas.

In this project the mapping was fixed and interpreted the same. A more adaptive mapping, perhaps that is introduced and based on simple principles, can work better. An open body plane is more, closed body is less. This can evolve over time, allowing the user and system to adapt. That could result in much more personalized and grounded in the users perception mapping that should produce a more clear mind model. That would also support the cultural backgrounds of the users and partner with them in the context of their own culture.

To summarize, the findings suggest that the clarity of the mapping and the intuitiveness of it is the factor that sets the system for success for this type of interaction.

7.0.2 RQ2

The AI was variably perceived as a partner in this experiment. Questionnaire responses showed that many participant felt like co-authors, but with some nuances. While P! empathized the AI's role in extending their own ideas ("filling in the blanks"), P2 acknowledged novelty but questioned the accuracy of the translation of emotion.

Surprise emerged also in the questionnaire where most participants report that it was pleasant. For p1, the system produced little novelty, while P2 highlights the AI's ability to generate phrasing beyond typical human expression. This aligns with previous work **??** and suggest that calibrated surprise is crucial. Meaning that too little reduces inspiration, but too much risks detachment.

Another crucial factor for the system is pacing and control. Questionnaire results showed mixed responses to the statement "I was able to guide the creative outcome through my body movements". P1 reported that the system felt "fast", limiting their sense of agency. This highlights even more the need for adjustable parameter as tempo and sensitivity. These findings suggest that design strategies for AI as a creative partner should combine:

1. Transparency of mapping: The user needs to understand their influence
2. Adjustable surprise: A way for the user to calibrate the AI divergence
3. Flexible pacing: Giving the user time to reflect and refine, control the session timeline

7.0.3 Ethical implications and societal impact

The development and deployment of an application like Pose Poet carry significant ethical implications and potential societal impact.

Privacy and surveillance risks in pose tracking While this project use a webcam-based pose estimation, none of the data associated with it was recorded or kept. However, this is still includes processing sensitive biometric data. This carries risks, especially if linked to an identity. Using sensitive data with complex algorithms could lead to lack of transparency and rise concerns about accountability and trust.

Bias in body language interpretation Ai systems, including those interpreting body language are susceptible to biases inherited from the training data. This can lead to misinterpretations or perpetuation of cultural stereotypes. In a project like this one, where bodily expression is being translated and connected to affective states and emotions algorithmic bias could lead to incongruence in the experience of the user. Moreover, hand gestures for example can have vastly different meaning across countries, cultures, and regions.

Authorship, intellectual property and value in art The co-creative nature of Pose poet raises complex questions regarding authorship, intellectual property, and the artist value of AI generated poetry.

In many parts of the world copy write laws are different but generally require "human authorship". As examined earlier in this paper, there is notion that protest against the co-collaborative notion of human and AI [25]. Arguing that AI's work is just human labor in disguise, could complicate the conversation around authorship. That is also perpetuated with many conversations around the data sets used to train generative AI. This raises significant IP infringements concerns and ethical questions about the use and compensation of original content creators.

Societal impact of embodied systems with generative Ai With the current widespread adoption of Ai in creative areas around us, a shift in educational focus could be needed. As traditional knowledge centered education has to adopt. Adaptation of new skills and more wide fields of expertise can be observed. A co-creative tool like Pose poet could be used to expose students to the world of poetry, Help students discover and bring the abstract concepts to more tangible interactions.

Accessibility and inclusion While this project focuses on embodied interaction as the primary means of creative engagement, it is important to acknowledge that a such design may unintentionally exclude users that are not fully body-abled. This concept assumes a baseline of a physical ability that not all users possess. For example, full-body participants while limiting opportunities for individuals with reduced mobility. HCI highlights that the design should ensure that the system are not only technicality robust but also socially inclusive.

The addaptive interaction mapping suggested in a earlier section could proof that the system will be able to produce results even with smaller scale movement or alternative means. Such approaches can examine eye gaze, limited hand movement and gestures. Introducing different modalities as discussed in an earlier section, could also allow for more diverse participation. That also suggest that "embodiment" does not necessarily mean also full-body as the theory suggests. Inclusion of non-body abled users also aligns with broader societal impact of AI in creative domains. Such systems should help and democratize the artistic expression, instead of introducing new boundaries.

8 Conclusion

Pose poet is a compelling new way of interaction between human movement and generative artificial intelligence. It offers a transformative way to engage with poetry and expression. The technical

feasibility of real-time pose estimation via webcam si well established with models like MeadiaPipe. The primary challenge lies in effectively bridging of the semantic gap between the raw kinesthetic data and abstract poetic concepts.

Pose poet is a successful attempt to try and bridge that gap and point the spotlight to the field. An application like that demands intuitive user experience and immersive design. Principles of embodied interaction and direct manipulation are crucial to empower users sense of creative control. iterative refinement of the output ensures that the human remains in control while partnering with the AI.

However, the development of a such system and the implications of it are not without its ethical complexities. Primary concerns over privacy, potential algorithmic bias in body language, the complex topic of authorship and exclusivity.

Looking ahead, pose poet framework opens several promising gaps for future research and development:

1. Enhanced semantic interpretation: Further research into more nuanced semantic mapping could be beneficial. Looking into dynamic pose sequences, abstract concepts and emotional states, could enrich a project like Pose poet.

2. Multi-modal generative AI: Exploring integration of additional modalities, such as voice tone or facial expressions as part of the input could enrich the output. Moving towards truly multi-modal AI system that more close resembles human communication.

3. Therapeutic and education applications: Investigating the potential pose-driven poetry generation in therapeutic rehabilitation, mental health, and educational settings, leveraging the form of self-expression and body awareness that is outside the mundane.

4. Ethical AI governance: Such projects can contribute to the enhancements of robust ethical guidelines and frameworks. Responsible innovating with embodied Ai and creative expression should central.

Pose poet stands in a crossroad between the virtual, where currently generative AI resides, and physical, where human embodied interaction and communication takes place. By navigating both worlds and complex challenges they entail, Pose poet seeks to bridge them. It serves as a meeting point and promises a new era of embodied poetics and experiences, where silent language of the body find the poetic voice of expression. All allowing deeper creative connections between humans and machines.

REFERENCES

[1] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, et al. Guidelines for human-ai interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*, pp. 1–13, 2019.

[2] A. Bandura. Social cognitive theory: An agentic perspective. *Annual review of psychology*, 52(1):1–26, 2001.

[3] A. Bryman. *Social research methods*. Oxford university press, 2016.

[4] S. M. Bsharat, A. Myrzakhan, and Z. Shen. Principled instructions are all you need for questioning llama-1/2, gpt-3.5/4. *arXiv preprint arXiv:2312.16171*, 2023.

[5] H. M. Cooper. Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in society*, 1(1):104, 1988.

[6] P. Cronin, F. Ryan, and M. Coughlan. Undertaking a literature review: a step-by-step approach. *British journal of nursing*, 17(1):38–43, 2008.

[7] M. Csikszentmihalyi, S. Abuhamdeh, and J. Nakamura. Flow. In *Flow and the foundations of positive psychology: The collected works of Mihaly Csikszentmihalyi*, pp. 227–238. Springer, 2014.

[8] P. Fagerberg, A. Ståhl, and K. Höök. Designing gestures for affective input: an analysis of shape, effort and valence. In *MUM 2003, Norrköping, Sweden*. ACM Publications, 2003.

[9] F. Giunchiglia, M. Bagchi, and X. Diao. Aligning visual and lexical semantics. In *International Conference on Information*, pp. 294–302. Springer, 2023.

[10] K. L. Hao-Chiang, M.-C. HSIEH, E. Z.-F. LIU, and T.-Y. CHUANG. Interacting with visual poems through ar-based digital artwork. *TOJET: The Turkish Online Journal of Educational Technology*, 11(1), 2012.

[11] C. Hart. *Doing a Literature Review: Releasing the Social Science Research Imagination*. Sage, Thousand Oaks, CA, 1998.

[12] K. Hook. *Designing with the body: Somaesthetic interaction design*. MIt Press, 2018.

[13] A. Kantosalo and S. Riihiaho. Experience evaluations for human–computer co-creative processes–planning and conducting an evaluation in practice. *Connection Science*, 31(1):60–81, 2019.

[14] A. Kantosalo and S. Riihiaho. Quantifying co-creative writing experiences. *Digital Creativity*, 30(1):23–38, 2019.

[15] S. Kulkarni, S. Deshmukh, F. Fernandes, A. Patil, and V. Jabade. Poseanalyser: A survey on human pose estimation. *SN Computer Science*, 4(2):136, 2023.

[16] Q. V. Liao, D. Gruen, and S. Miller. Questioning the ai: informing design practices for explainable ai user experiences. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pp. 1–15, 2020.

[17] L. Loke and T. Robertson. Studies of dancers: Moving from experience to interaction design. *International Journal of Design*, 2010.

[18] A. Menon. *From Mind to Machine: An Embodied Approach to Image Creation with Generative AI*. PhD thesis, OCAD University, 2024.

[19] D. Mochon, M. I. Norton, and D. Ariely. Bolstering and restoring feelings of competence via the ikea effect. *International journal of research in marketing*, 29(4):363–369, 2012.

[20] M. Nayak and K. Narayan. Strengths and weaknesses of online surveys. *technology*, 6(7):0837–2405053138, 2019.

[21] J. Nielsen. 10 usability heuristics for user interface design. `https://www.nngroup.com/articles/ten-usability-heuristics/`, 1994.

[22] M. I. Norton, D. Mochon, and D. Ariely. The ikea effect: When labor leads to love. *Journal of consumer psychology*, 22(3):453–460, 2012.

[23] B. Ollila and J. Jantas. The definition of poetry. *New York: Teaching Matters Inc*, 2006.

[24] J. Randolph. A guide to writing the dissertation literature review. *Practical assessment, research, and evaluation*, 14(1), 2009.

[25] A. Sarkar. Enough with "human-ai collaboration". In *Extended abstracts of the 2023 CHI conference on human factors in computing systems*, pp. 1–8, 2023.

[26] B. Shneiderman. Direct manipulation: A step beyond programming languages. *Computer*, 16(08):57–69, 1983.

[27] B. Shneiderman. Creativity support tools: accelerating discovery and innovation. *Communications of the ACM*, 50(12):20–32, 2007.

[28] C. Teddlie and A. Tasshakori. A general typology of research designs featuring mixed methods. *Research in the Schools*, 13(1), 2006.

[29] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, et al. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*, 2022.

[30] J. D. Weisz, J. He, M. Muller, G. Hoefer, R. Miles, and W. Geyer. Design principles for generative ai applications. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pp. 1–22, 2024.

[31] C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz, and M. Shah. Deep learning-based human pose estimation: A survey. *ACM computing surveys*, 56(1):1–37, 2023.

CONTENTS

1 Introduction **1**

1.1 Problem statements 1

1.1.1 Research questions 1

1.2 Thesis overview 1

2 Theoretical background **1**
2.1 Mixed-methods research 1
2.2 Human-centered design 1

3 Literature review **1**
3.1 Literature review 1
3.1.1 Search process 2
3.2 Embodied interaction 2
3.3 UX within AI generated content 2
3.4 Creative agency 3

4 Methodology **3**
4.1 Participants 3
4.2 Research design and materials 3
4.2.1 Pre-task questionnaire 3
4.2.2 Demographics 3
4.2.3 Previous experience with AI 3
4.2.4 AI-collaborative session 4
4.2.5 Technical setup 4
4.2.6 Materials 4
4.2.7 Post-task questionnaire 4
4.2.8 Post-task interview 4
4.3 Procedure 4
4.4 Data analysis plan 5
4.4.1 Data from pre-questionnaire 5
4.4.2 Data from the the post-experience questionnaire 5
4.4.3 Data from interview 5

5 Application design **5**
5.1 Overview of Human pose estimation 5
5.1.1 Technology 5
5.1.2 The semantic gap 6
5.1.3 Affective feature extraction and interpretation 6
5.2 Human-centered design in practice 6
5.3 AI prompt construction 6
5.3.1 Model 6
5.3.2 Prompt 6
5.4 Application workflow 7
5.5 UI Pilot Study 7
5.6 User Experience and Interface Design of Pose Poets 7
5.6.1 Visibility of System Status 7
5.6.2 Supporting Mental Models 8
5.6.3 Direct and Intuitive Interaction 8
5.6.4 Aesthetic and Minimalist Design 8
5.6.5 Summary 8

6 Results **8**
6.1 Data sets 8
6.1.1 Participants 8
6.1.2 Pre- questioner 9
6.1.3 Post-questionnaire 9
6.1.4 Semi-structured interviews 9
6.1.5 Summery 9

7 Discussion **9**
7.0.1 RQ1 9
7.0.2 RQ2 10
7.0.3 Ethical implications and societal impact . . 10

8 Conclusion **10**