

# Intentional Spaces

As AI understands more of the physical world, it enables us to make sense of our surroundings in new and easier ways.

Kay van den Aker

Spring 2024  
MFA Interaction Design  
Umeå Institute of Design



# Abstract

Even though most of our life happens in the physical world, most applications of AI are digital. We have to explain our intention and context to chatbots for them to support us with things happening in the world around us.

Archetype AI, the collaboration partner of this thesis, is developing a Large Behavior Model (LBM) that perceives and reasons about the physical world in real time by fusing multimodal sensor data and natural language. This thesis is a side track to the technical development of this LBM, exploring the landscape of interactions enabled when AI understands the world around us.

A series of interventions with physical AI as a design material is presented to illustrate what the relation with such intelligent systems could look like in the future. The research also highlights the evolving role of designers, which transitions from crafting explicit interactions to shaping the behaviour of intelligent and context-aware environments.

Lastly, the research raises questions to start asking as AI, and its role in our lives is evolving.

*Thanks to Leonardo, Marta, Alex, Marie, Regina, Emile,  
Lesley, Lukas and everyone who supported and inspired  
me throughout this thesis. It means a lot :)*

# Table of Contents

Abstract	2
Table of Contents	4
Introduction	5
The rise of AI	6
Archetype AI	7
In short	7
Background	8
From Information to Intelligence	9
Intentional Spaces	9
Post-Phenomenology	10
Approach	11
Physical AI as a material	12
Stage 1 – Understanding the material	12
Stage 2 – Working with the material	12
Interacting with the material	13
Stage 1 – Understanding the Material	14
MediaPipe	15
Combining Models	16
Haptic Gloves	17
Making sense of sensors	18
Reflection: Context is everything	19
Stage 2 – Working with the material	20
Stethoscope	21
Binoculars	23
Reflection: AI as an Agent	25
Sketch	26
Dynamic Actuators	28
Reflection: Experience the invisible	32
AI as a 6th sense	34
The context is the prompt	36
Analysis	38
Discussion	39
Conclusion	41
References	42
Appendices	43
Appendix A – Dynamic Actuators	43

# Introduction

What is physical AI, and where did it come from.

# The rise of AI

In today's world, humans are surrounded by countless tools and technologies designed to enhance our capabilities — cutlery, printers, glasses, and phones, to name a few. These artefacts, shaped by us to improve our lives, have in return shaped our environment and how we interact with it (Elhacham et al., 2020). With the rapid evolution of technology, especially lately in artificial intelligence (AI), our relationship with tools, and the world around us is evolving at a tremendous speed.

AI, such as ChatGPT, has shown a remarkable capacity for handling complex tasks at unparalleled speeds, suggesting its potential to tackle a wide range of challenges.

*“Looking at LLMs as chatbots is the same as looking at early computers as calculators. We're seeing an emergence of a whole new computing paradigm, and it is very early.”  
(Karpathy, 2023)*

As of now, AI still mostly remains contained in the digital world, transforming digital inputs into digital outputs using vast datasets. Similarly in the past with powerful new technologies, like the computer, when we haven't defined fundamentals for it yet (the archetypes), we start with simple text-based interfaces like the terminal.

*“But if you need a manual, it's not ‘natural language’ anymore. Once you start talking about ‘prompt engineering’, you're describing command lines - you're describing what came before GUIs, not what comes after them.” (Evans, 2023)*

Similar to how the internet started to show its full potential when devices like the smartphone allowed us to seamlessly access the web from virtually anywhere, looking something up online became a standard extension of human's capability. AI is a similar platform.

*Technology (AI) itself is not what's changing our daily lives, rather its embodiment into something that reshapes our interactions with the world around us, and with each other.*

# Archetype AI

OpenAI's GPT-4 and Google's Gemini are now starting to integrate data from the web and the built-in sensors of the device they operate on, like webcams. Starting to connect AI with the world around it.

At this forefront, Archetype AI<sup>1</sup>, a startup originating from Google's ATAP department, leads the way in developing a physical AI<sup>2</sup> platform called a Large Behavior Model (LBM). This LBM makes sense of the physical world around it by looking for behavioural patterns through various sensors, such as cameras, distance sensors, and in particular, radar. Archetype AI is developing a radar together with Infineon that can detect even the smallest kind of motion, like whether someone is breathing in the room (Maxfield, 2023).

*“The combination of Archetype AI’s AI ... with Infineon’s 3D radar sensors will allow developers to uncover the hidden patterns of behavior in unstructured sensor data and blend real-time sensor data with natural language to create a living view of the world to analyze complex sensor-based data.” (Maxfield, 2023)*

So through various sensors, the LBM creates a comprehensive and nuanced understanding of what’s going on in a space, in real time. It knows what people are doing, what objects moved, what happened, etcetera.

Currently, one can only interact with this LBM through a terminal interface, asking for example “what happened at X” or “show me when you detected Y”. But there’s a whole new field of interactions, experiences and relations that are being enabled by such systems, and exactly that forms the focus of this thesis.

The research was executed in collaboration with Archetype AI, aiming for a well rounded blend of academic, industry, technical and philosophical perspectives.

## In short

This thesis digs into what physical AI as a material means for the world of interaction design. It aims to explore the possibilities and opportunities that arise when a space understands, reasons, and interacts with the people and things inside of it. How our experiences with our environment change, and our relationships with technology transform through physical AI.

*How can, and should designers use physical AI as a material?*

The intended audience for this research, as of now, are designers and technologists — people who are shaping how we engage with technology in the future. In a longer term, the understanding of the material aims to empower designers to shape meaningful experiences in contexts where physical AI proves relevant.

---

<sup>1</sup> <https://www.archetypeai.io/>

<sup>2</sup> Physical AI refers to artificial intelligence systems that process and analyse information from one or more sensors in the physical world in real time (e.g. camera, radar, IMU, microphone).

# Background

The new medium needs new metaphors.

# From Information to Intelligence

With the rise of intelligent systems, we can express our intention directly, instead of executing a set of fixed interactions. Instead of doing 6 clicks, 2 swipes and a scroll to play a song, we can just say “play X”.

With an intelligent system, the more it understands our context and intentionality, the more complex tasks it can do with less complex interactions.

*“Turn on the lights in the living room”*  
*“Turn on the lights” – knowing the user’s location*  
*Turns on the lights, recognising the user does not want to be in the dark*

The main interactions we will still have with technology will be negotiating intentionality.

## Intentional Spaces

Imagine a future where the full space is an agent, and the AI is an autonomous, social, reactive and pro-active entity that co-exists and interacts with humans in a space. In this research such spaces are referred to as Intentional Spaces.

We can move away from terminal looking interfaces, and beyond voice interfaces. We can express our intention in whatever way suitable for that context, and the AI could make sense of it. The AI could reason about what action through what actuator is appropriate to respond, and then do that action. The human perceives this, and reasons if that was “success” or if something else needs to be expressed.

The same way the human takes initiative, the AI can also do that if it sees the need. To warn the user about danger for example, or support when needed. Once the AI has agency to be proactive, there will be a true dialogue to negotiate the shared intentionality between the human and the AI. And the stage of interaction to support that dialogue can be directly in our physical environment. Negotiation should move from terminal looking interfaces into the physical world.

*The new medium needs new metaphors.*

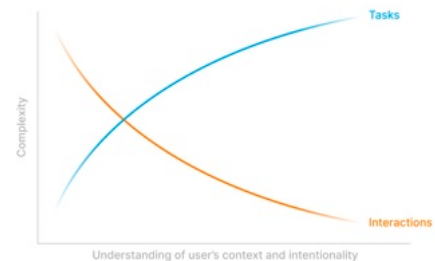


Figure 1. Complexity relation

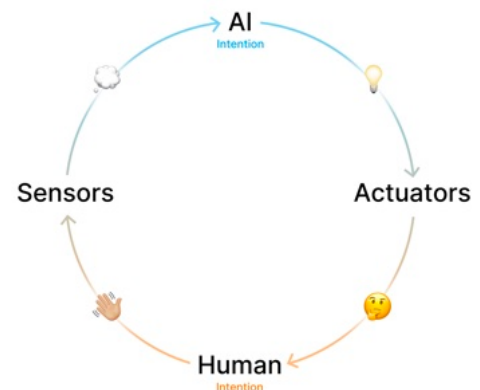


Figure 2. Negotiating intentionality

# Post-Phenomenology

The philosophy of post-phenomenology is used as an inspirational lens throughout the thesis. Taking its definitions for the different human-technology relations as a framework to identify, understand and explore the experiences that come into existence using physical AI.

Post-phenomenology is a philosophical and theoretical framework exploring the various types of relationships between humans, technology and the world. Post-phenomenology acknowledges the importance of technology in shaping human perception and experience, emphasising that technology is not just a tool but also a mediator of our interactions with the world.

An extremely simplified overview of the types of human-tech relations is provided below, the arrows indicate intentionality.

Relation	Diagram	Explanation
Embodiment	(human - technology) → world	Technologies which reshape the user's experience, with the device itself in some ways taken into the user's bodily awareness. (e.g. glasses)
Hermeneutic	human → (technology – world)	Technologies which are used through an act of perceiving and interpreting the device's readout. (e.g. a wristwatch)
Alterity	human → technology (world)	Devices to which we relate in a manner somewhat similar to how we interact with other human beings. (e.g. ATM displaying questions)
Background	human → (technology / world)	Technologies that make up the user's environmental context. (e.g. heating system)
Cyborg	(human / technology) → world	Technologies that blur the boundaries between body and artefact. (e.g. neuro implants)

**Figure 3.** This table is based on the author's interpretation of 'A Field Guide to Postphenomenology'. (Rosenberger et al., 2015)

# Approach

*How to make sense of physical AI as a design material?*

# Physical AI as a material

As physical AI is being developed now, the research aims to get a thorough understanding of its capabilities, and what kind of interactions and experiences it could enable. Approaching physical AI as a design material. (Redström, 2005)

*“Emerging approaches to design such as interaction, experience and service design, often utilising new technologies with almost ‘immaterial’ properties, seem to point to a situation where the material ‘thing’ as we used to know it is replaced by communication, information, systems and infrastructures.” (Redström, 2005)*

The research aims to approach exploring the materiality of physical AI by using a research-through-design methodology (Koskinen et al., 2011), iteratively thinking through making. Exploring the boundaries and opportunities of the material through a series of sacrificial prototypes and interventions (Chung, 2016). Exploring and experiencing possible futures, aiming to frame how designers of future experiences can make sense of, and use this material. These prototypes also serve as platforms for discussion, enabling to easily discuss the implications of the material.

*A series of interventions with physical AI as a design material.*

The goal of approaching the research from a research-through-design perspective, combined with post-phenomenology and the technical expertise of Archetype AI, is to frame the material with nuance. Zooming in with the interventions, and zooming out through the philosophy aims to facilitate a comprehensive understanding, considering both human and technology.

The first questions to answer through the research are:

- How can the stage of interacting with AI move from a screen, to the world around us?
- How can humans and AI do something together in the physical world?
- What could those relations look like when AI can adapt to our context, considering where we are, what we’re doing, with whom, how much time we have, how we’re feeling etcetera.

*Explore the landscape of interactions enabled by physical AI*

## Stage 1 – Understanding the material

Defining the primitives, principles, opportunities and boundaries of the material is an important starting point to begin exploring use cases, but also for others to get a grip on what it means to work with physical AI.

Similar to how many people now have a basic understanding of the boundaries of voice assistants, the first part of the research aims to provide such a mental model for physical AI.

## Stage 2 – Working with the material

After the dialogue with the physical AI material, making sense of it and understanding its capabilities in the different types of human-technology relations — this thesis aims to provide a horizon of potential applications and design opportunities. Communicating some of those contexts through high fidelity prototyping or video.

## Interacting with the material

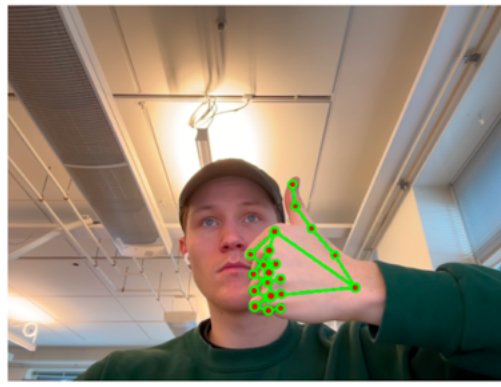
To design interactions and relations with this new material, inspiration and approaches will be derived from academic research such as *Designing Behavior in Interaction* (Ross et al., 2010) and *Aesthetics of Being Together* (Kuenen, 2018). These studies offer perspectives and insights into designing for technology that mediates our interactions and experiences with the world around us, and the other humans in it.

*“As design moves into the realm of intelligent products and systems, interactive product behavior becomes an ever more prominent aspect of design, raising the question of how to design the aesthetics of such interactive behavior.” (Ross et al., 2010)*

*“An aesthetic of being together is a proposition of a different fundament for interaction design practice. In addition to referring to properties of things and qualities of interacting with things, it refers to the kind of relations that come to expression between people interacting with each other with these things.” (Kuenen, 2018)*

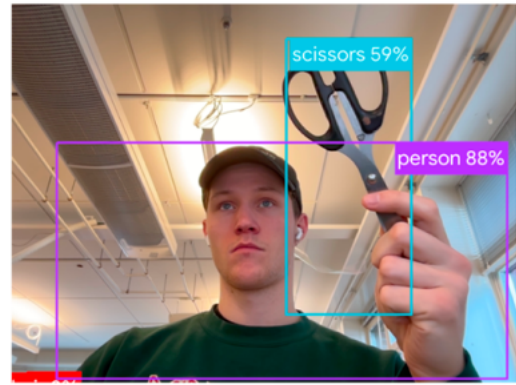
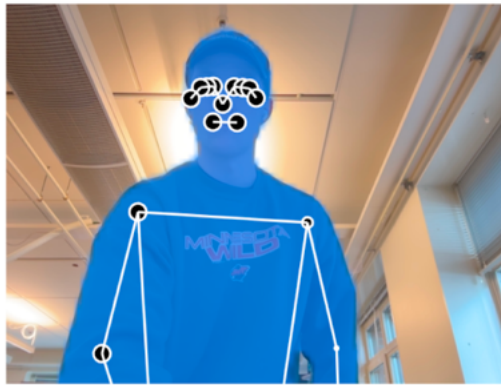
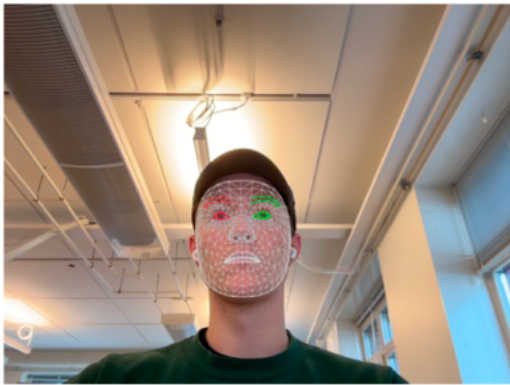
## **Stage 1 – Understanding the Material**

Gain understanding of the landscape of Artificial Intelligence (AI), Large Language Models (LLM) and Machine Learning (ML).



Inference time (ms): 21.1  
Thumb\_Up 58%

Inference time (ms): 16.1  
water bottle 18%  
beaker 14%  
remote control 14%



## MediaPipe

### Goal

Understanding the capabilities of existing machine learning models.

### Prototype

Various MediaPipe examples made by Google were tried.<sup>3</sup>

### Insights

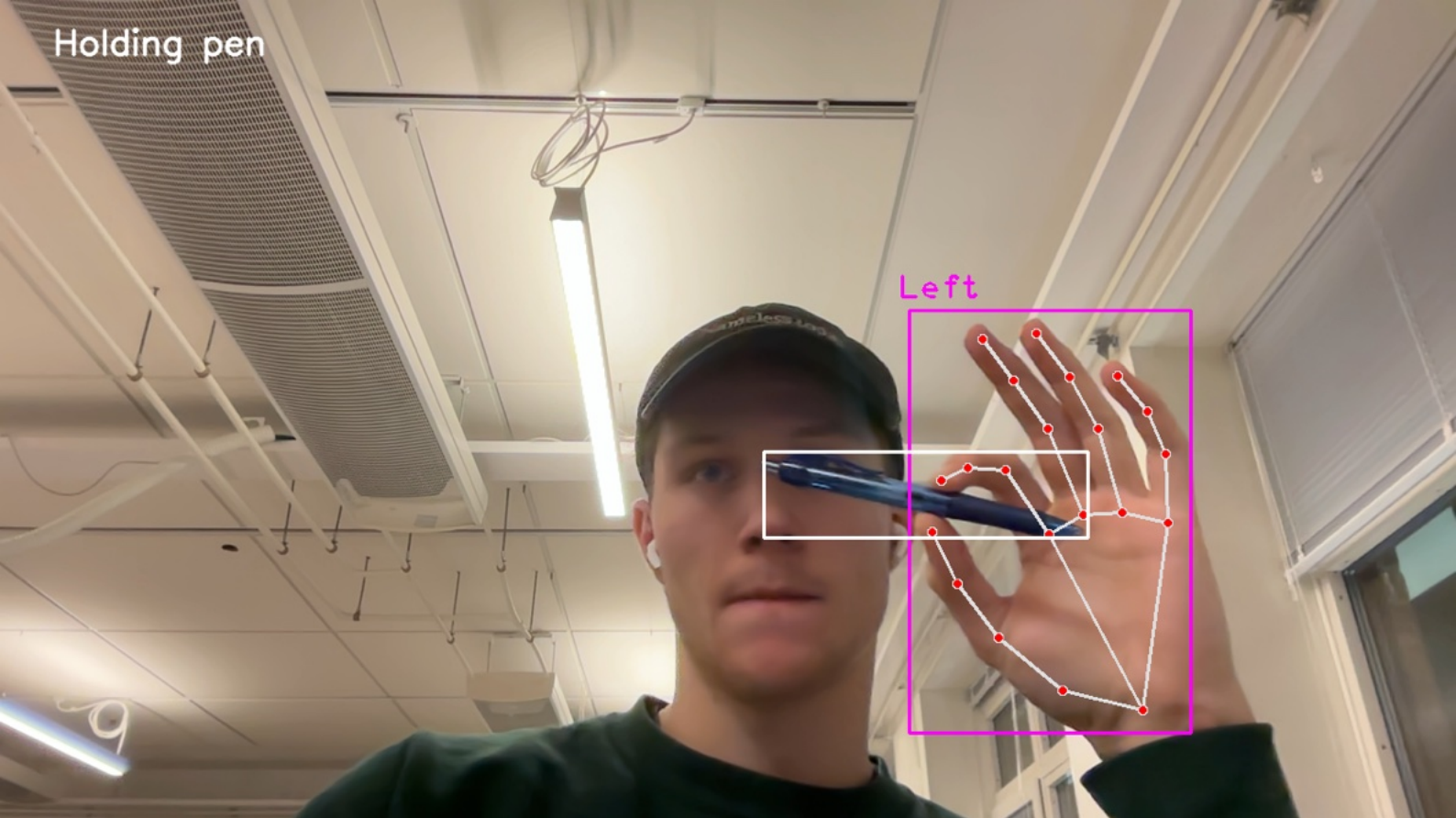
It's often misunderstood what the differences and capabilities are of "AI", "LLM" and "ML". AI is the broader concept of machines simulating human intelligence, ML is a subset of AI focused on algorithms learning from data, and LLMs are a specific type of ML model specialized in understanding and generating human language. Each model has a specific purpose, and one cannot, and should not, just 'use AI' for everything.<sup>4</sup>

*Designers should not look at models in silos, but rather at glueing the right models together, leveraging their different properties depending on the context at hand.*

<sup>3</sup> <https://mediapipe-studio.webapps.google.com/home>

<sup>4</sup> Considering the tech available at spring 2024

Holding pen



## Combining Models

### Goal

Explore possibilities and opportunities of combining multiple models.

### Prototype

MediaPipe's hand tracking was combined with a Zero-Shot Object Detection model (OwlViT). Tasking it to find if the bounding box of the hand is overlapping the prompt to look for.

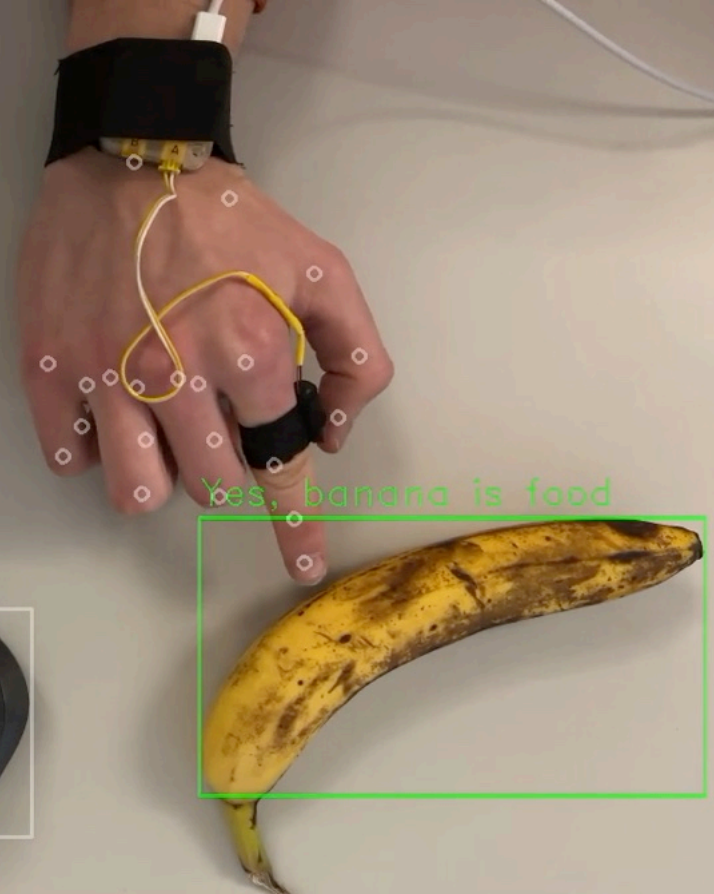
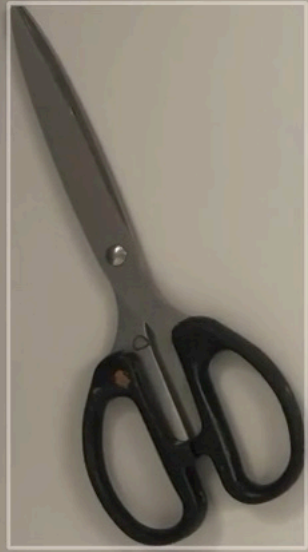
```
prompt = "a photo of a pen"
```

### Insights

The MediaPipe hand tracking can run with unnoticeable lag, however prompting things to look out for in a camera data stream using Zero-Shot Object Detection is quite processing heavy, and resulted in around 1 FPS on a M2 MacBook Air. Pre-trained models to identify predefined objects are much faster and can run with unnoticeable lag.

*Models that do less complex tasks are obviously faster, aim for the leanest model for your purpose to optimise performance and user experience.*

Is this food?



## Haptic Gloves

### Goal

Exploring how technology could express itself towards humans through haptics.

### Prototype

A pre-trained object detection model (YOLO v8) was tasked to show if the object that is overlapping with the location of the index finger (MediaPipe) is food. The model expresses itself to the human through haptics, incorrect is a quick double pulse, correct is a slower single pulse.

### Insights

Moving feedback away from the screen onto our body has great potential to make technology a 6th sense. This example to identify food is oversimplified and could be easily done by most people, but perhaps in the future the technology could reason about more complex things and guide us in new and easier ways.

*AI has the capability to translate one sense (vision) into another one (touch), which opens up many interesting use cases, especially considering accessibility contexts.*



## Making sense of sensors

### Goal

Make sense of sensor data without pre-training, using an LLM.

### Prototype

A simple script was created that recorded accelerometer data while the human presses the button on the microcontroller. Upon release, it would send the recorded data to OpenAI's LLM and task it to reason about the movement of the accelerometer. It quickly became clear that these tasks are quite abstract if the model has no idea what even is up. Multiple prompts were tried, and providing it with example positions drastically increased the accuracy.

```
prompt = "What movement is in this handheld accelerometer data stream? Each reading is 'X,Y,Z', the measurement interval is 50ms, the unit is m/s^2. Device on a flat surface measures '0,0,10'. Focus on the major movement, ignore small movements. Answer in one sentence, include the movement, the direction and the speed. The data set:" + data
```

### Insights

The exploration showed an interesting interaction paradigm of explaining states to the technology, imagine having a simple accelerometer sensor, glue it to a door, close the door, explain it "this is closed", open the door, explain it "this is open". And then the model could reason about the state of the door, even if the data does not exactly match the predefined "open" or "closed" state.

*The more an LLM understands the meaning and context of the data its looking at, the more accurately it can reason about it.*

# Reflection: Context is everything

Each model has its own quality, and it should be carefully considered what kind is suitable for the task at hand. See the figure on the right for the type of models available at May 2024 on [huggingface.co](https://huggingface.co), one of the biggest collections of open-source generative AI models.

It became clear that Large Language Models are very capable of looking at sensor data and describing it. However context is key, the more information the model has about the context of the data, the more detailed output it can provide.

The new landscape of interactions enabled by (a combination of) those models forms the starting point of the next section – working with the material.

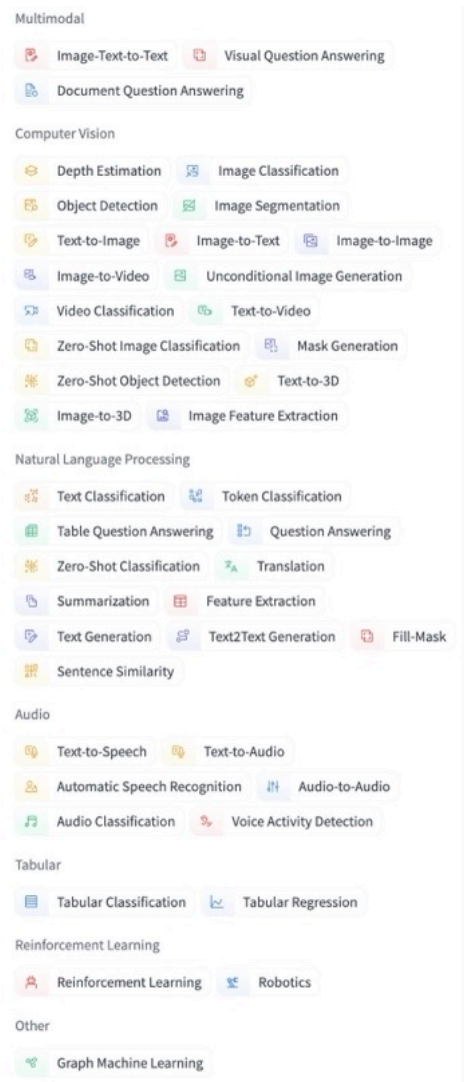
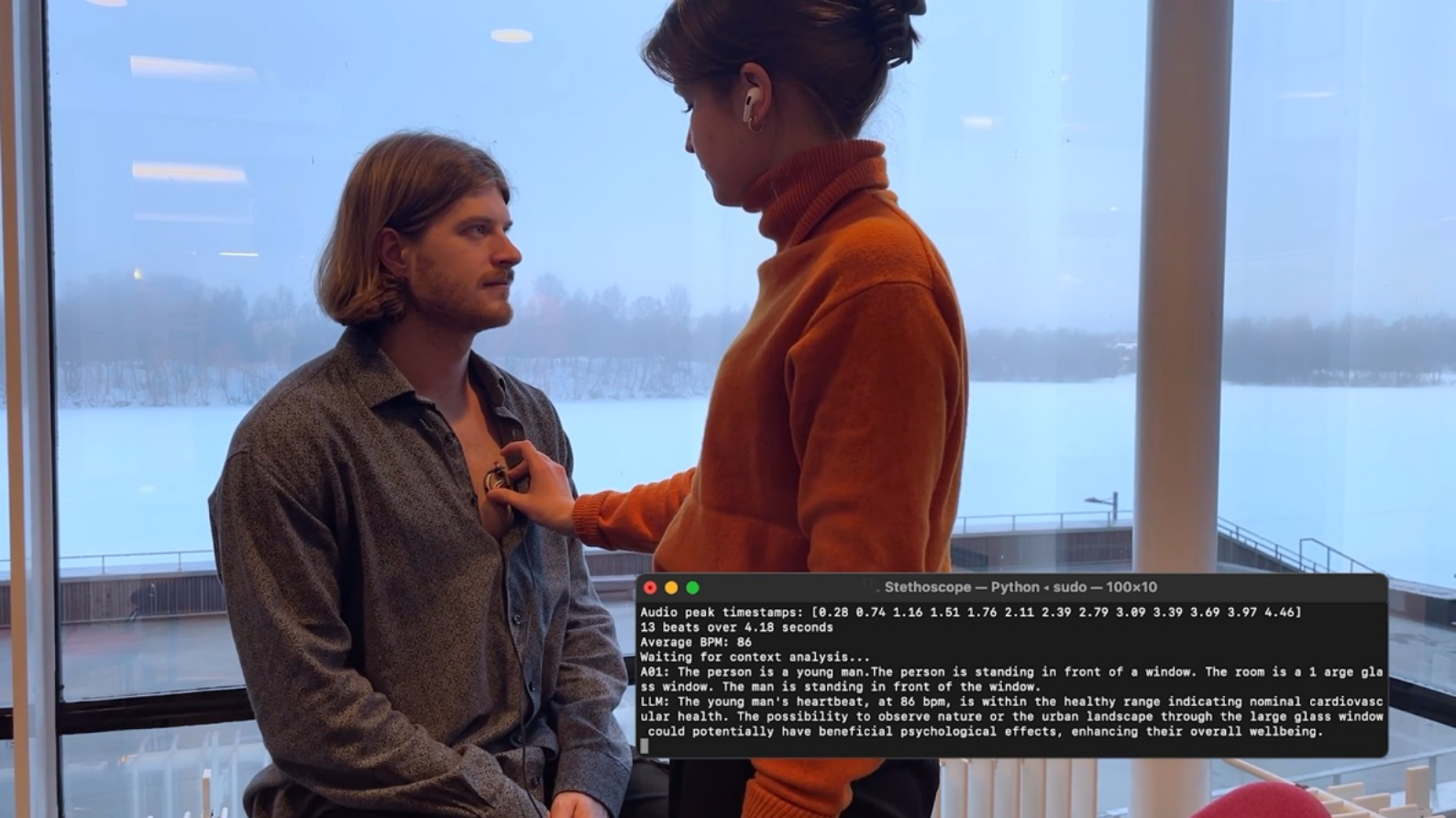


Figure 4. Type of models available at [huggingface.co](https://huggingface.co) in May 2024

## **Stage 2 – Working with the material**

Interventions applying physical AI in various contexts, gaining insights into working with it as a design material.



## Stethoscope

### Goal

Explore augmenting a human sense (hearing) with insights of AI. What could it feel like to listen to a human body with a stethoscope, and simultaneously hear AI's understanding of it.

### Prototype

The stethoscope consists of a microphone, a push button on top and headphones. By default the microphone recording is playing on the headphones of the doctor to hear whatever they're listening to. If the button is pressed, it captures a picture of the patient and starts listening to the microphone for 5 seconds. It uses the captured picture to analyse the person and its context using Archetype AI's Summarize API.

```
context = summarize("Describe the age of the person") + summarize("What is the person doing?") + summarize("Describe the room")
```

This would return for example "The person is a young man. The person is standing in front of a window. The room is a large glass window. The man is standing in front of the window.". It became clear that splitting prompts into more specific questions worked much better than asking generic questions. Targeting specific things such as the age of the person, the activity and the room.

After the 5 seconds of recording, a Python library called Librosa looks for the peaks in the recorded audio file and determines a BPM.<sup>5</sup> It then feeds that BPM together with the context analysis to OpenAI's LLM.

```
prompt = "Pretend you are a talking stethoscope, the information you have is:" +  
context + " The person has a bpm of " + bpm + ". Answer with only a short answer  
with insights about the health of the person, including the bpm and the context.  
Don't introduce yourself."
```

The result of this prompt is then turned into speech using OpenAI's Text To Speech API and played on the headphones of the doctor.

## Insights

It was interesting to see how this prototype was perceived by some as a 'low-tech' solution even though the tech is quite complex. It might suggest that moving away from screens gives people the impression that things are more simple. It was also interesting to see that the doctor could maintain eye contact and focus on the patient, stay more present. Similarly, asking the patient for silence when it wants to listen to the heart rate, it could nudge the patient when the AI was speaking.

*The fact that the AI's insights were only audible to the doctor was good, as then they could combine it with their own understanding, insights and reasoning before acting. Empowering humans, keeping them at the centre, instead of replacing them.*

Some notable responses were when it explained "a cap is not influencing the heart rate, but the ceiling fan might influence the climate which could influence it". Things that are super obvious for us, such as the cap, might not be for the LLM.

---

<sup>5</sup> This is far from perfect but works fine for the scope of the prototype.



## Binoculars

### Goal

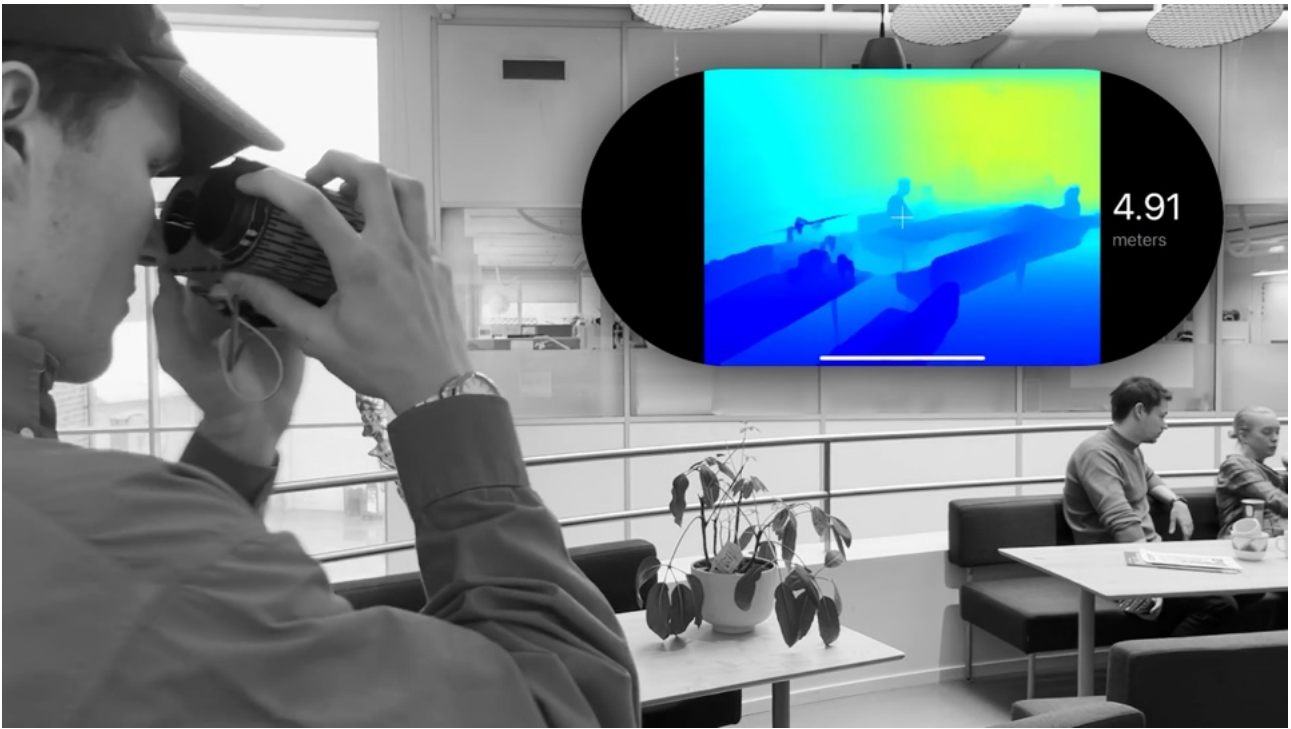
Look through the eyes of the AI and gain understanding of how it makes sense of the world.

### Prototype

A binoculars was created using an iPhone, a push button and a rotary encoder. A SwiftUI application was made, using the iPhone's LiDAR sensor to get the distance at the centre of the camera. The user could switch between colour and depth view by rotating the rotary encoder, and get an "AI analysis" by pressing the button on top. This would send the image to Archetype AI's Summarize API and task it to describe the object in the centre of the image.

```
prompt = "What object is at the centre of the picture?"
```

This would return text, and a bounding box of that object. It then calculates the center of that bounding box and uses the LiDAR data to get the distance to that object. The bounding box in combination with the distance was then used to calculate the dimensions of the object. Crude, but sufficient for the scope of this exploration.



## Insights

It was interesting to see how the prototype sparked so much curiosity with people while the demo was crafter. A depth camera in itself does not really show us things we “can’t sense”, but the way information is presented with the colorful view alters our perception of reality at that moment. We know the table is that far, and the plant is that far, but we rarely look at it like this.

*Combining the LiDAR with the AI analysis gave people a tool to look at the world through the eyes of the AI, seeing how it interprets the context.*

# Reflection: AI as an Agent

Throughout the first explorations it became clear that there was a lack of agency from the AI, it was more of an instrument with dedicated purpose. The focus and goal were defined very explicitly, instead of dynamically adjusting, which is a core quality of AI.

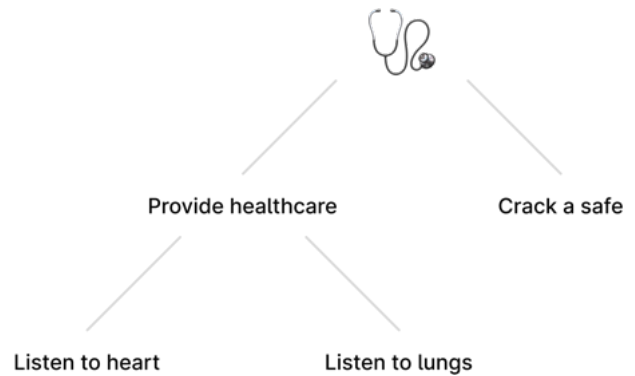
## Introducing Lenses

A sensor is basically a static piece of tech that outputs data. But what if we can set a lens, and adjust how AI interprets this data, what it should look out for. What the user is interested in, in that specific moment. In a way tune the AI to match the intention, and similarly it could suggest a “lens” that it thinks is relevant.

Similar to how a camera lens filter does not alter the world around us, it just alters the way someone perceives it. Highlighting different things depending on what the user is aiming to see. With AI this could also be suggesting a lens to show something it thinks could be interesting.

*The lens will define the way the AI will try to make sense of the data, but how do we set this lens?*

For example with the stethoscope, the initial goal is “hearing heartbeat”, but what if the user could zoom out and make it “listening to a body”, or even more to “providing healthcare”, or tune it to a different focus, “listening to breathing” or a totally different one like “cracking a safe” or “translate Spanish”. And ideally the AI would also proactively suggest things it could do in that moment, continuously negotiating the shared intentionality.

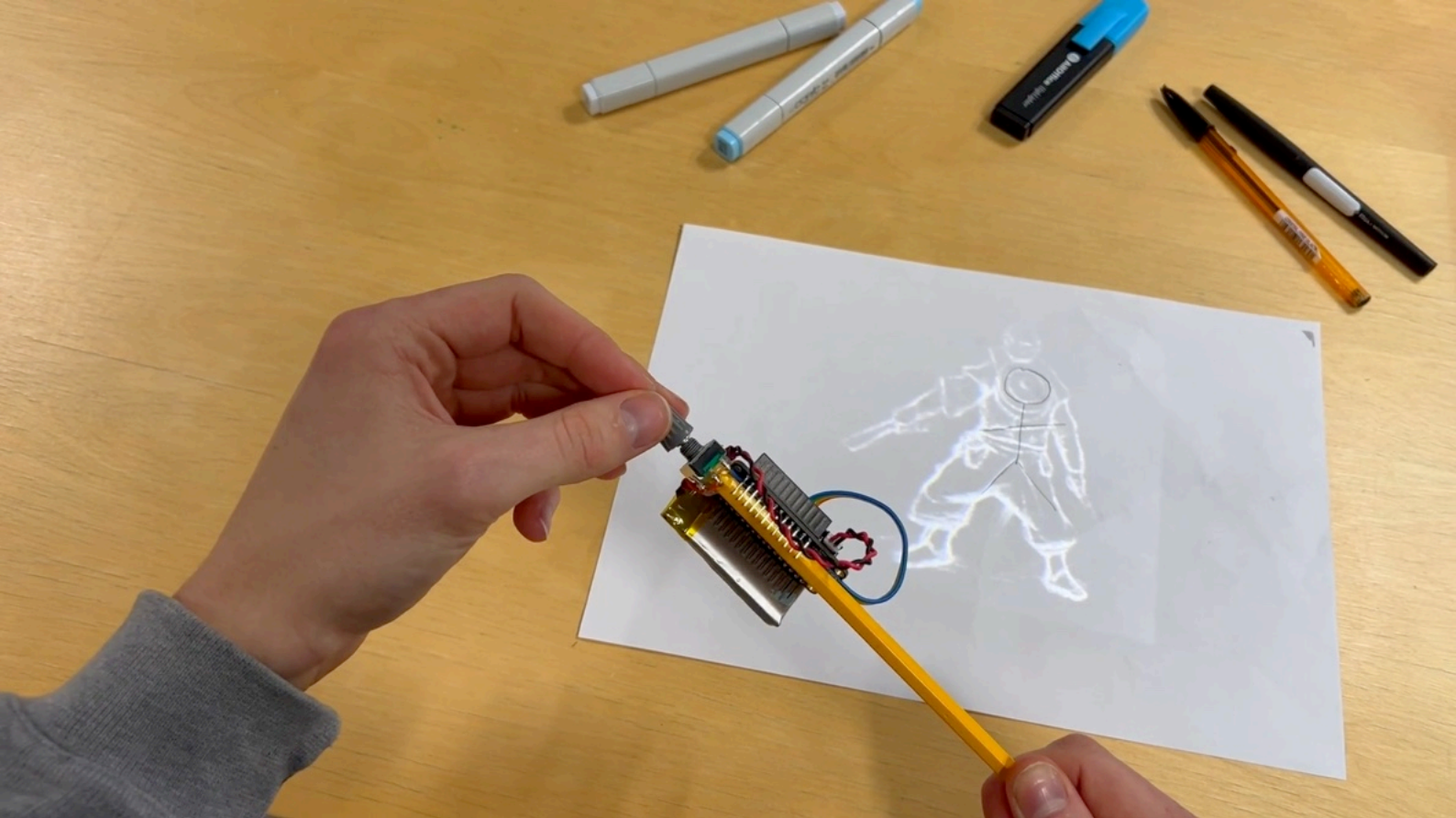


Likewise with the binoculars, what if the user could set a focus like “plants”, “humans” or “furniture”. Or set it to “animals” while in a zoo, and it would look for animals and tell the user about them. With a contextual understanding, the AI could also do that proactively and suggest “animals” as a lens because it senses the user is in a zoo.



## Next steps

The upcoming activities will try to focus more on giving agency to the AI, moving away from the “AI as an instrument” towards “AI as an agent”. Aiming to have a more outside-in approach, instead of inside-out. Not trying to embody AI in a tool, but rather as the space, as defined in the original brief. Prototype experiences that bring more agency to the tech.



## Sketch

### Goal

Explore how AI could be less instrumental, and take more agency to inspire the user.

### Prototype

A camera captures a sketch and uses that together with a prompt as input for a Latent Consistency Model to generate images.

```
prompt = "A hand drawn sketch"
```

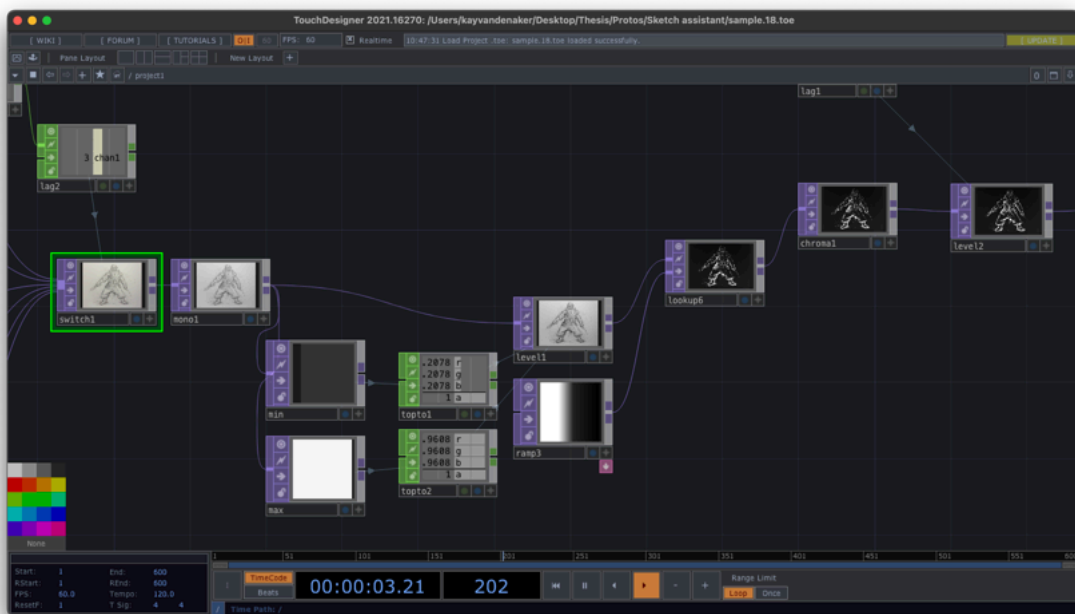


After playing around with the prompt, adding more details about the background gave more consistent results.

```
prompt = "A hand drawn sketch with a white background"
```



These generated images are then adjusted in TouchDesigner to turn them into the white sketches with black backgrounds, which are then projected back onto the paper. When trying to generate white sketches on black backgrounds, the results were much less accurate, so this adjustment was done in post. Presumably cause the pencil sketch style is more prominent in the training data.



The pencil is modified to allow for interactions with the generated content. The rotary encoder on top of the pencil controls the TouchDesigner script through Bluetooth. Rotating it fades in- and out the generated overlays, and clicking it gives a new one.

## Insights

It was interesting to experience using traditional physical tools with familiar gestures, but augment them with AI and see how it changes the way we could sketch.

*As the prompt was open ended, the AI would generate a sketch of whatever it saw in the input, this randomness showed users things they didn't think of themselves.*

This was inspiring and showed things the humans were not thinking about at all. It was also the first prototype that showed some sort of 'entering' into our physical world, instead of being an instrument where the intentionality only comes from the user.

The functionality was still very defined, and the question arose how we could negotiate with it. Nudge it into certain a direction and have a dialogue with the space instead of having fixed input and output.



## Dynamic Actuators

### Goal

Give the AI a body to express itself in the physical world.

### Prototype

A Python script prompts the user with the message it should communicate to the subject.

```
message = input("Enter the prompt: ")
```

After the user enters the message, an image is captured to make sense of the context and the subject using Archetype's Summarise API. Two separate specific prompts are used, as this often resulted in more detailed output, compared to one longer or more abstract prompt.

```
context = summarize("Describe the age of the person") + " " + summarize("What activity is the person doing?")
```

The message and the context are then combined in a prompt to OpenAI's LLM, which is tasked to create a JSON object. The JSON is forwarded to an Arduino that controls 4 different actuators; a servo arm, a speaker, a lamp and/or a fan.

```
prompt = """"You are an intelligent assistant designed to interact with a person in a
room. """" + context + """". You have 4 options to communicate with the person: making
a wave with a little arm on the table the user is sitting on (the arm is a bit noisy
when it moves), beeping a speaker, blinking an RGB LED in a big desk lamp, or
blowing a small fan towards the person. Your task is to support the person by using
these actuators to communicate effectively. Craft a JSON response to convey the
message of: """" + message + """"'.
```

Here are the instructions for the control:

```
[
  {
    "actuator": "servo",
    "waveSpeed": VALUE_FROM_0_to_100,
    "waveDuration": VALUE_IN_SECONDS,
    "waveAngleStart": ANGLE_FROM_0_TO_90,
    "waveAngleEnd": ANGLE_FROM_90_TO_180
  }, {
    "actuator": "speaker",
    "beepFrequency": VALUE_FROM_30_TO_20000HZ,
    "beepVolume": VALUE_IN_DECIBEL,
    "beepInterval": VALUE_IN_SECONDS (0 is continuously on),
    "beepDuration": VALUE_IN_SECONDS
  }, {
    "actuator": "led",
    "colorR": VALUE_FROM_0_TO_255,
    "colorG": VALUE_FROM_0_TO_255,
    "colorB": VALUE_FROM_0_TO_255,
    "blinkInterval": VALUE_IN_SECONDS (0 is continuously on),
    "blinkDuration": VALUE_IN_SECONDS
  }, {
    "actuator": "fan",
    "fanDuration": VALUE_IN_SECONDS
  }
]
```

```
Send a JSON with a relevant response to communicate with the person. Be mindful to
be non-invasive, only if really needed. The JSON can include one or multiple
actuators."""
```

The LLM will reason about a suitable interactions considering the message, the subject and the context. It tries to understand what the prompt means for the subject, and how can it express that through the actuators available.

## Example

In the image shown on the previous page, the message was:

```
"There's a snake in the cup"
```

The image captured by the webcam resulted in the following context analysis:

```
"The person is a young man. The person is standing in front of a window with a glass
door."
```

The LLM then reasons about this message in relation to the subject and context and returns:

“The LED is chosen to grab the person's attention, and the color red is used to signal danger. The sound is then employed to emphasize the urgency of the situation, and the specific frequency is set to be within the range of the human voice to simulate urgency and alarm. The wave produced by the arm is adapted to simulate the slithering movement of a snake.”

The LLM seems to understand that a snake is dangerous for the young man. It flashes the LED with the color red to signal danger. It also reasons that sound should be used to emphasise the urgency of the situation, and a frequency is chosen in human voice range to simulate urgency. It even came up with a metaphor to use the waving arm to simulate the slithering movement of a snake.

See Appendix A for more examples.

## Insights

LLMs have the capability to reason about the context and what interactions are suitable for that context. It can reason whether it should act in the background or the foreground. Moving from focused interactions in our centre of attention, through periphery to implicit interactions outside our attentional field. In that sense moving across the whole interaction attention continuum (Bakker et al., 2016) as the AI sees fit.

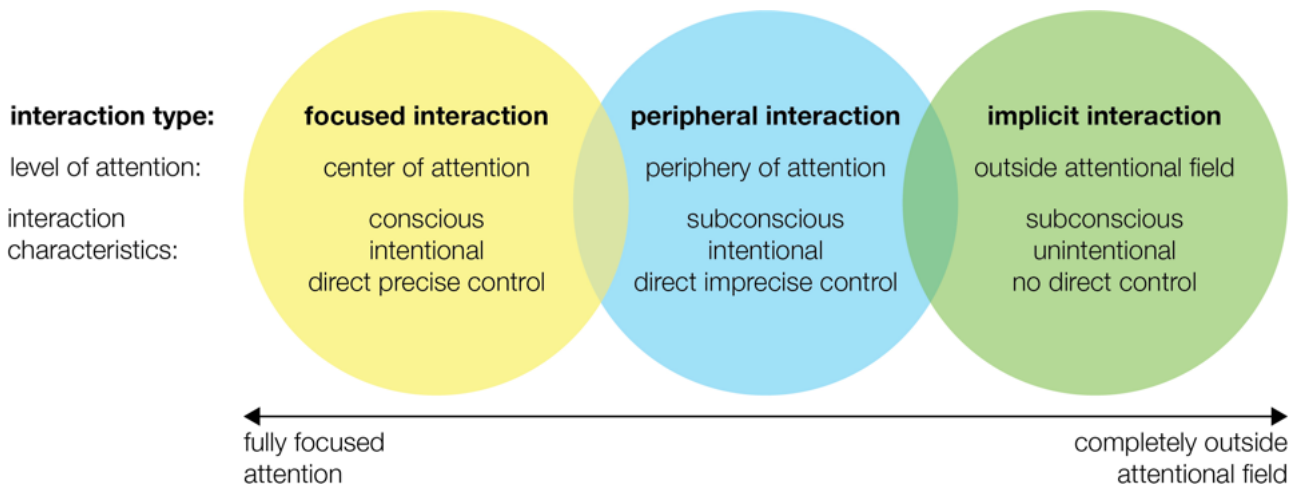


Figure 5. The interaction-attention continuum.

The intelligence allows for the interactions to become very fluid. Compared to the rather static relations defined in post-phenomenology, this technology can move into whatever shape it has access to and see fit for. We are not only expressing our intention to technology, it also starts to express it back, for example when Spotify or Netflix expresses what it thinks we want to see by the means of showing that on the home page.

Humans set tasks and desired outcomes together with AI, then AI determines appropriate interactions through its sensors and actuators. Humans intervene when it's inappropriate, and the AI learns. Humans can negotiate their shared intentionality with the AI. The interactions and interventions to do that don't happen on a screen, but directly in the physical world around us. <sup>6</sup>

*When a complete space is sensing and actuating through physical AI, it creates the notion of a space as an agent – the AI becomes an autonomous, social, reactive and pro-active entity that co-exists and interacts with humans in the space.*

---

<sup>6</sup> Assuming sufficient sensors and actuators are present to enable this.

## Reflection: Experience the invisible

The series of edited images are aimed to explore what different gestures could look like to negotiate shared intentionality with the space.

The user can explicitly share their intention by saying it out loud (assuming there is a microphone). But instead the user does a gesture with both hands pointing towards the context of the lens they're aiming to set. The AI will then suggest lenses considering the subject, objects and context and communicate that through the actuators available.



Figure 6. Requesting a lens, and requesting a new lens

After the human and the AI aligned on the lens, it can move to our periphery in the form of for example haptics. The bandwidth of the actuator can be much lower when the human knows what the AI is trying to express.



Figure 7. Feeling the invisible

For example, a vibration might be unclear when it comes out of nowhere, but when the user knows it is trying to communicate whether the plant needs to be watered or not, it is enough to understand if the haptics is a yes or a no.

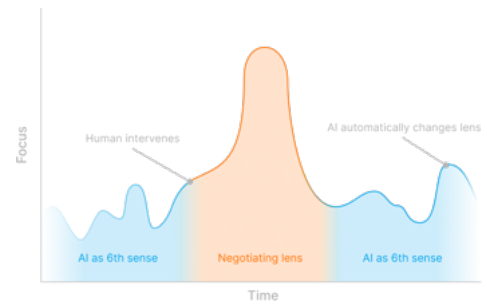
We can start to see how this relates back to the interaction attention continuum (Bakker et al., 2016), negotiating the lens likely takes some focus, we might talk, do a gesture or touch something. But then once the AI understands the intention of the person, it sets a suitable lens and shows the person the relevant information through whatever actuator available and suitable.

In a way it's like the extradiegetic narrator lens in movies, a voice that narrates what we see on the screen. This voice knows what's happening outside of the frame. In this case, we can get an understanding of what's happening outside our frame of perception.

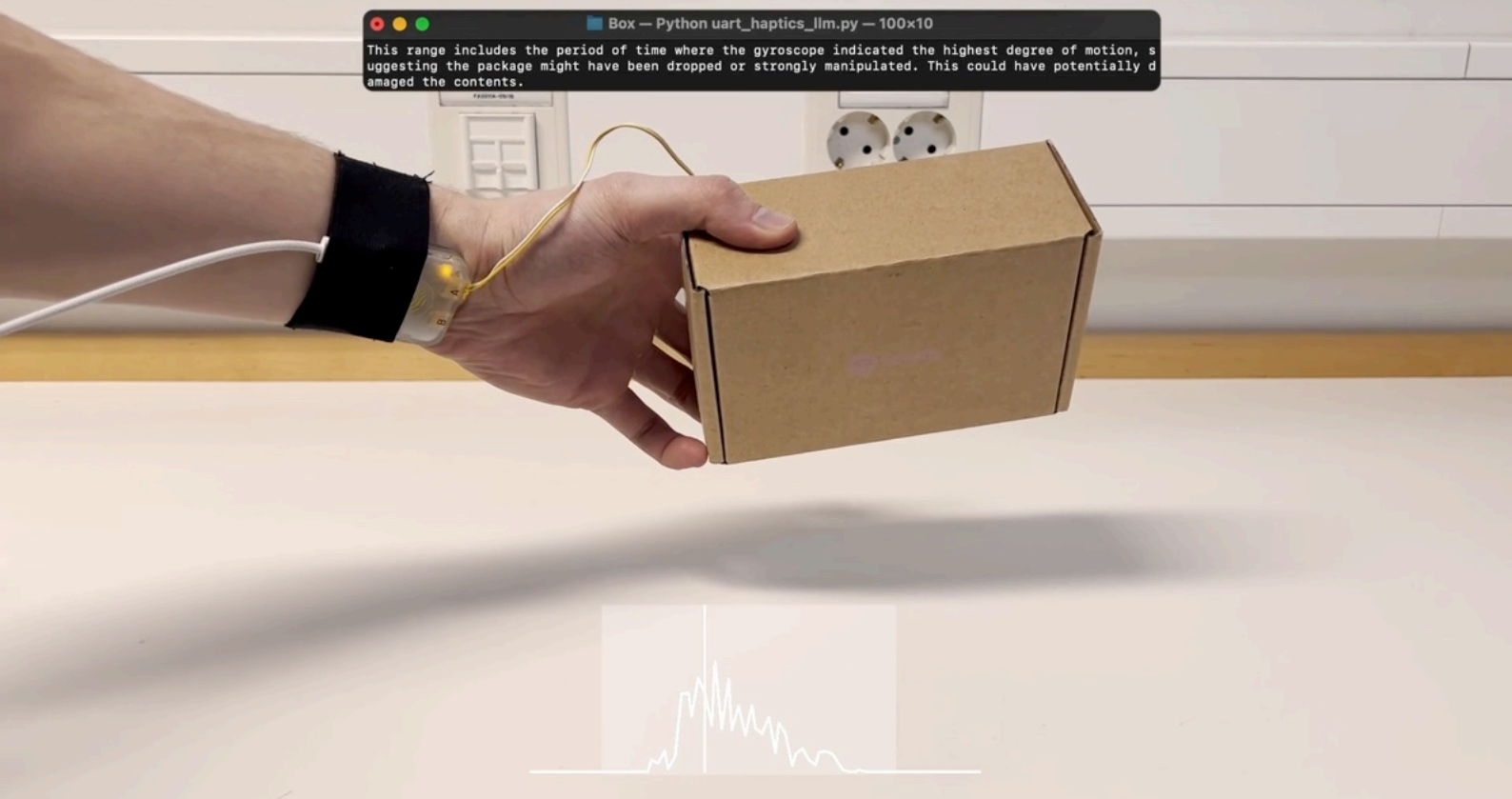
*We can sense what happened before in a space, what might happen next, and sense things we cannot with just our naked bodies. The role of the AI here is to translate its understanding of the physical world into something relevant to the users context in a way that they can understand.*

Assuming the physical AI has a thorough enough data set that allows it to sense and support with what is relevant, and hallucinations are minimised so the humans can 'rely' on the AI.

*The AI shifts between being an agent and being an instrument.*



**Figure 8.** Focussing on negotiating a lens, and AI as a 6th sense.



## AI as a 6th sense

### Goal

Explore how AI can be embodied as a 6th sense, giving a person a sense of whether a package might be damaged or not.

### Prototype

The gyroscope data of the package is stored, when the user picks up the package, the LLM looks at the data and picks the most relevant data to show to the person through haptics, to give them an idea of how the package was handled, and whether it's damaged or not. The prompt shows how the intent of the human is hard coded.

```
prompt = """You are an intelligent assistant designed to interact with a person through a haptic bracelet. You have the gyroscope data of a small cardboard box parcel. The data is as follows:""" + data + """". The interval between each data point is 100ms. You want to select a range of the gyroscope data by giving the start and end index of them in the array. This will then be played back to the person when grabbing the package through the haptic bracelet, that person wants to feel if the parcel might be damaged. 0 in the gyro data corresponds to the haptic vibrator being off, and 20 means it's totally on. Craft a JSON response to showing the most important range, including an explanation what you think might have happened to the package.
```

Here are the instructions for the control:

```
[{
    "explanation": "EXPLANATION_OF WHY THIS IS THE MOST IMPORTANT_RANGE",
    "indexStart": INDEX_START,"indexEnd": INDEX_END
}]
```

```
Reply with only a piece of JSON including the relevant response. Make sure to use the structure provided."""
```

So the LLM shows a moment of the objects motion history that it thinks is relevant for that person, and then communicates through what actuator is available, the haptic bracelet.

Some results:

- This range has been selected because it includes the highest gyroscope readings, indicating potentially the most significant movements or shocks experienced by the package. This could imply potential damage.
- This range contains drastic changes in the gyroscope's data, indicating intense movement. It may suggest that the package fell down or experienced a similar substantial jolt, which could have triggered potential damage.
- This range shows where there was significant movement or shock to the package. This could indicate that the package was dropped or handled roughly, and may possibly be damaged.
- This range of the gyroscope data reveals that there was a significant shake or shock event, likely from the parcel being dropped or mishandled in some way, possibly leading to its content being damaged.

## Insights

It's interesting how the AI can move to the background, outside of our focus, with these low bandwidth haptic actuators. Once the AI understands what the person is wanting to do in their environment (know whether a package is damaged or not), the lens is set, and it can communicate a much more specific message through lower bandwidth actuators, like a 6th sense.

The understanding of the space does not only allow for 'smart' objects to become more intelligent, it also allows new interactions with 'dumb' objects. For example a syringe in a medical context, imagine a doctor can sense what liquids have been in the syringe before, the nurse can sense where to use it, and the cleaning staff can sense where to throw it away.

*People have different relations to different objects in different contexts, and what data is relevant for them is therefore also super dynamic.*

Or in this example, the 'dumb' box all of a sudden has many more properties that the human can sense, beyond the physical/spatial properties of the cardboard box and its contents.

*Without changing a single thing about an object, a whole new (personalised) set of interactions to make sense of them in their context comes into existence.*



## The context is the prompt

### Goal

Explore how the physical context can be used as the prompt, and how shifting focus towards an actuator could express the need for assistance on that actuator.

### Example

In this example, the doctor examines the patient's knee, the AI comprehends this intricate context and generates insights on demand based on all available data – for example medical records, history and scans.

The doctor can express its need for assistance by looking at the screen, suggesting he wants the AI to show visual support on that actuator. Subtle facial expressions can then be used to get new suggestions, or enlarge them.

### Prototype

The prototype uses MediaPipe Face Landmarkers in a Node application running in the browser. It detects the presence and orientation of a face. The markers are used to recognise facial expression to generate new (hard-coded) insights, or enlarge them.

## Insights

Normally, a concise LLM prompt looks something like this

```
system prompt = capabilities + context + expected data + goal
```

So what if the capabilities and context are known by the physical AI, and the goal and expected data is reasoned about through the sensors. Then the user can express their intention to the system and get help dynamically. Like in the example, the doctor simply looks at the screen (signalling it wants to see something there – the expected data) while inspecting someone's knee (context). Through the context and capabilities that the physical AI has, it can reason about how to best help the doctor with providing healthcare (goal), and provides some order scan of the patients knee with its thoughts augmented on top.

Emphasising that the doctor is still combining this with its own judgement, not replacing but empowering the people in the situation.

What if looking at the screen is already the prompt? What if your context is the prompt?

```
System Prompt = capabilities + context + expected data + goal
```

- Capabilities = X
- Context = sensed in digital and physical environment (doctor X is inspecting knee of patient Y)
- Expected data = expressed by user (doctor looking at screen, expecting something to be shown there)
- Goal = supporting the doctor with providing accurate healthcare, in this case about the knee (understood from context)

It's interesting how moving our focus towards an actuator can be used as an input to evoke an action. It's easy to express towards actuators that are spatially anchored like visuals, haptics etc, but harder with ambient actuators like sound, focussing towards a speaker might be weird when it's all around us. You cannot 'listen towards something', not the same way we can look at something.

The doctor could also keep his hands on the knee while looking at the screen, not physically having to move towards the screen. Allowing for a seamless support from the AI.

*When parts of our bodies are occupied in activities in the physical world, other parts can still be used to express our intention towards our environment. AI can reason, learn and make sense of those expressions to support accordingly. Which can be unique for each context.*

# Analysis

*So, what does it mean to use physical AI as a design material?*

## **From Information to Intelligence**

Different sensors, enable different interactions to express our intention towards the physical AI. The most straight forward is typing, but as the examples in this thesis showed, there are many opportunities to interact with AI directly in our physical environment.

The point is that the physical AI lowers the barrier for us to access information, and make sense of our environment. At the same time it also makes sense of how to respond accordingly with the actuators it has available. So the interactions are extremely context dependant.

## **Intentional Spaces**

Once the AI is at the stage where it becomes an autonomous, social, reactive and pro-active entity that co-exists and interacts with humans in the space. It means that not only we can summon it as an instrument, it can also proactively enter our lives when it sees the need for it, in case of an emergency for example. AI can reason about how much focus it should take from the human in the specific context to support in the right way. It's important for designers to find the right balance between controlling the AI, and leaning back and letting it happen.

The research suggests Intentional Spaces could support us in two main ways; amplify our cognitive strengths and extend our physical capabilities. It can help us to enhance our understanding of the world and simplify making sense of our environment, instead of taking over and doing the things for us.

## **Post-Phenomenology**

The role of physical AI can be very dynamic, so the relations move across the relations defined in post-phenomenology, and from agent to instrument, and from background to foreground.

Sometimes the user might want to embody Ai's understanding of our environment as a 6th sense, feeling on a haptic bracelet wether a plant needs watering through haptics for example. Because when physical AI tracks what's happening to all objects in a space, everything comes alive around us. We could all of a sudden ask a syringe what has happened to it before, or where it needs to go, without changing a single thing about the syringe itself. So our relations to other objects and technologies in a space could all change through physical AI.

# Discussion

*Things to consider when designing with physical AI.*

## **Products designed to be inside Intentional Spaces**

Physical AI enables a whole new field of products to immerse, tailored specifically to be used in intentional spaces. A simple example could be a drill with a simple red/green LED, which could indicate if the drilling location is correct or not. As the intelligence lives outside of the products, they can simply connect to the physical AI, and display its insights through simple actuators.

## **Connecting to other humans**

An interesting notion immerses when the AI senses how a human does a certain craft in a space, to support another human to do the same. Imagine for example how an expert surgeon on the other side of the world could do a surgery, the physical AI observes closely, and can then support a surgeon on the other side of the world to do it in a same way.

Or how a painter held a brush. The input for these learnings could also be existing media such as videos. What if it can show you how to paint based on a series of Bob Ross videos.

## **We are still animals**

Unlike most technologies, we are organic beings. We have hormones, moods and emotions. This makes us unique, and is a huge quality of being a human compared to machines. We are creative and care for people and things. In those different states we also express ourselves differently to our environment, also depending on who we are with, and where we are etcetera. AI would ideally learn and adapt to this as well, and respond accordingly to a sad, stressed or relaxed mental states. But the complexity of human behaviour should not be underestimated.

## **Cultural biases**

In the Dynamic Actuators chapter, we saw that the AI chose red to signal danger. Using red for danger is not a universal thing for all humans though, in Chinese culture red represents luck and happiness. This is a simple example, but cultural values and biases reflected in the data set can get extremely complex and should be carefully considered.

More details about the subject or context are needed for the AI to reason about what values and culture fit the person, but the complexity of this should not be underestimated.

## **Character of a space**

Any intelligent system has a certain personality, even if it doesn't have one programmed, it has a way of expressing itself that we interpret in a certain way. The character of an intentional space should not always be the same. Maybe in a home context it can be more warm, in a construction site more stoic, or in a hospital more direct. It should match the expectations of the people inside the space.

## **Role of interaction designer**

In the context of physical AI, the role of designers will shift from crafting explicit interactions to shaping the behaviour and character of intentional spaces. For example in the Dynamic Actuators exploration (page X), it is hard coded for the AI to “Be mindful to be non-invasive.”. This works for the context it was functioning in, but sometimes it should maybe not behave like that. Similar to the Making sense of sensors exploration (page X), one example is provided to the LLM as a reference. Similarly, examples of expected behaviours will be needed for the AI in more complex scenarios, so it can reason about how to act appropriately.

In a construction context for example, should the AI tell the boss whether the workers were doing their job or not. Whose side should it chose? Interaction designers should carefully consider the behaviour, especially when the AI is having a mediating role between people.

## **Ethics**

Some possible applications of physical AI could end up in Wall-E like scenarios.

*“Things, both emerging and already-established comprehensive intelligent systems, gradually turn our world into a “new uterus”—warm and soft, compliant, in accordance with our preferences, providing, shielding, and never irritating. Is this the brave new world, a cozy technological womb of Eden?” – Bruno Gransche*

So it’s important to ask ourselves the question “when is physical AI relevant?”. I believe, generally speaking, when something’s at stake beyond frustration or discomfort. For example, danger prevention on construction sites or supporting emergency room personnel. Such contexts could also justify the privacy aspect and initial investment to set up a physical AI platform.

## **AI is not perfect**

Most of this thesis assumes that AI will get to a point where it can comprehend a thorough level of complexity, but doing this accurately is a huge challenge. It will take some time before it does not hallucinate anymore to an extend we can actually rely on it.

## **When is physical AI relevant**

Overall, it’s extremely complex to make claims about where AI might serve a meaningful purpose. There needs to be further investigation in specific contexts to judge its use and ethical implications.

# Conclusion

*As AI understands more of the physical world, it enables us to make sense of our surroundings in new and easier ways.*

The series of interventions present a glimpse into the landscape of interactions enabled by physical AI. What it could be like to co-exist with an autonomous, social, reactive and pro-active entity that interacts with us in the physical world.

The insights gained from working with physical AI as a design material form the outcome of this thesis.

On an interaction design level, the project explores how the dynamic relation between human and technology could look like in the future.

On a technical level, the project outlines possibilities and challenges that arise when incorporating physical AI into the products, services and systems around us.

# References

*OpenAI's ChatGPT software has been used throughout the writing of this report for refining purposes.*

OpenAI (2024). ChatGPT (Accessed 2024-05-22) [Large language model]. <https://chat.openai.com/chat>

## Academic references

Bakker, S. & Niemantsverdriet, K. (2016). The interaction-attention continuum: considering various levels of human attention in interaction design. *International Journal of Design* 10(2): pp. 1-14.

Chung, C.M. (2016). Sacrificial Concepts. <https://medium.com/@christinamchung/sacrificial-concepts-200993246364> (Accessed 2024-05-22).

Elhacham, E. et al. (2020). Global human-made mass exceeds all living biomass. *Nature* 588(7838): pp. 442-444. <https://doi.org/10.1038/s41586-020-3010-5>

Evans, B. (2023). Unbundling AI. <https://www.ben-evans.com/benedictevans/2023/10/5/unbundling-ai> (Accessed 2024-05-22).

Karpathy, A. (2023). <https://twitter.com/karpathy/status/1707437820045062561?lang=en> (Accessed 2024-05-22).

Kuenen, S. (2018). Aesthetics of Being Together. <http://umu.diva-portal.org/smash/get/diva2:1256460/FULLTEXT01.pdf> (Accessed 2024-05-22).

Koskinen, I.K. et al. (2011). Design research through practice: from the lab, field, and showroom. Morgan Kaufmann Publishers, Inc. <https://doi.org/10.1016/B978-0-12-385502-2.00015-8>

Maxfield, M. (2023). Infineon and Archetype to Develop AI Sensor Chipsets. <https://www.eejournal.com/article/infineon-and-archetype-to-develop-ai-sensor-chipsets/> (Accessed 2024-05-22).

Redström, J. (2005). On technology as material in design. *Design Philosophy Papers* 3(2): pp. 39-54. <https://doi.org/10.2752/144871305X13966254124275>

Rosenberger, R. & Verbeek, P.P.C.C. (2015). A field guide to postphenomenology. In: Rosenberger, R. & Verbeek, P.P. (eds.) *Postphenomenological Investigations: Essays on Human-Technology Relations*. Lexington Books, pp. 9-41.

Ross, P.R. & Wensveen, S.A.G. (2010). Designing behavior in interaction: using aesthetic experience as a mechanism for design. *International Journal of Design* 4(2): pp. 3-13.

Xu, L. et al. (2023). Soli-enabled noncontact heart rate detection for sleep and meditation tracking. *Scientific Reports* 13: 18008. <https://doi.org/10.1038/s41598-023-44714-2>

# Appendices

## Appendix A – Dynamic Actuators

These are the results of various tests with the dynamic actuators, it shows interesting reasoning from the LLM about the prompts. The headings are the prompts, the bullets are responses, the blue text is added to elaborate and highlight some of the results.

### It's windy outside

- Fan will be used to simulate the sensation of wind. The blowing speed will be set moderately to depict wind outside and not a storm.
- The blowing of the small fan can replicate the feeling of wind to signify that it is windy outside.
- Using the fan to emulate the feeling of wind to indicate that it is windy outside. Also utilizing the speaker to create a windy sound effect for extra emphasis.
- To imply that it's windy outside, naturally the best approach is to use the fan to simulate the wind. A gentle wave from the servo also adds a visual hint, like tree branches swaying in the wind.
  - *Using the servo arm as a metaphor for tree branches swaying in the wind, very creative*
- The fan will be used to represent the wind outside. The servo's wave and the LED's blink have been added to draw the user's attention to the fan's action.
- We will use the fan to communicate the windiness. Sudden wind caused by the fan will best represent actual wind which might be experienced outside.
- Using the fan to create a physical sensation of wind, along with a slow wave of the servo arm to simulate the waving of tree branches in the wind. This should communicate the concept of 'windy'.
- The fan is used to simulate the wind, giving a realistic sense of the outside weather condition. The LED light is set to a cool blue color, which could resemble a windy day's sky, and it slowly blinks to get user's attention without startling.
  - *Being cautious as this is not urgent*

### I need to go to bed soon

- I chose to blink the desk lamp slowly to simulate the winding down of the day and the approaching bedtime. I chose a comforting and warm orange color to set the mood and to not harm the eyes, as bright or white lights can wake the person up instead. Additionally, a subtle breeze from the fan can stimulate the feeling of cool night air, which can aid in making the person feel drowsy.
- I chose the LED actuator to blink slowly, simulating getting tired and going to sleep, and the speaker to play a soft and low-frequency beep, mimicking a lullaby or calming noise.
  - *Mimicking a lullaby and slowly blinking to wind down*
- First, I will wave using the servo to gently catch the individual's attention. Then, I will use the LED to blink at a slow pace and keeping the light dim to simulate the effect of a setting at a slow pace simulate sun, which signifies it's nearing bedtime. After this, I will use the fan to blow a gentle breeze to create a soothing environment that indicates it's time to relax and prepare for bed.

## The building is on fire

- The situation is very serious and demands immediate actions. Therefore all actuators are used. The waving arm attracts visual attention, speaker is used to give an auditory signal, the LED blinks red as a universal sign of danger, and the fan is used to possibly make the person feel the heat.
- An urgent but non-invasive method to communicate the urgency of the situation is to wave the arm rapidly, create a continuous and loud beep, blink the led in a rapid pace with the color red and and loud beep, in a rapid pace running the fan for several seconds. These gestures have to be significant, immediate and unmistakable.
  - *The last word is very interesting for urgency; unmistakable.*
- Given the severity and urgency of the situation, all actuators are going to be used. The speaker is used to create a loud sound attracting immediate attention with determining frequency and multiple harmonics. The RGB LED light will rapidly blink in bright red light, a universal signal for danger or alarm. The fan will blow to further draw attention to our signals, and finally, the servo's arm will wave to create a physical presence of danger.
- The alarm will be sounded using the speaker at maximum intensity to quickly grab attention. The LED will flash red to indicate danger. To ensure the message is understood, the servo arm will continuously wave to attract attention and the fan will blow as a physical indication of urgency.
- Given the urgent nature of the message it's important to capture the person's attention immediately. The LED should flash red to symbolize danger, while the speaker should make a high-pitched and loud sound, to mimic an alarm. The arm should wave rapidly and the fan should blow to signify the intensity.
  - *It seems to reason about the relation between urgency and high pitched sounds, flashing red.*
- Chose the arm waving at maximum speed and with a wide range to grab attention, the fan to give physical sensation of urgency, and the speaker making a loud recurring beep, to indicate an emergency. The LED is blinking with red color as it is typically associated with danger.
- To convey the urgency and gravity of the situation, all four actuators will be used. The waving arm will provide a physical cue of intensity, the speaker will offer a high pitch sound to signal immediate attention, the LED will blink rapidly in red, symbolizing danger, and the fan will blow to hint the presence of smoke.
  - *Another metaphor, the fan to hint the presence of smoke.*
- In case of an emergency like a fire, the signal must be hard to ignore. Thus, I'll use all available actuators to ensure the user notices. The servo will wave to attract attention; the speaker will generate an alarm sound; the LED will blink red, a common color associated with danger; and the fan will blow to simulate the sensation of heat.

## There's a snake in my cup

- The wave is chosen to first grab the person's attention, then a fast beep is sounded as an alert. The lamp is set to blink red signalling danger. The fan is used last, to blow away from the person as if to suggest moving away from the table.
- To convey urgency, the speaker is used to produce a fast-paced beeping, the servo makes a frantic wave, the LED blinks red rapidly, and fans blow for short time for tactile sensation.
- Use arm to grab immediate attention, blink LED in red color to deliver a sense of urgency, and beep speaker with increasing sound frequency to raise alarm. The sound chosen is square due to its harshness, making it more alarming.
  - *The LLM reasons here about harshness of the square waveform, making it more alarming.*
- The LED is chosen to grab the person's attention, and the color red is used to signal danger. The sound is then employed to emphasize the urgency of the situation, and the specific

frequency is set to be within the range of the human voice to simulate urgency and alarm. The wave produced by the arm is adapted to simulate the slithering movement of a snake.

- *Using the movement of the arm to simulate the slithering movement of a snake.*

### **The person hasn't taken a break in a while**

- As the person is busy taking a selfie, we opted for using the fan as it is less invasive and does not interfere with the person's activity. Also, a gentle breeze from the fan may symbolize a breath of fresh air which can be associated with taking a break. Additionally, good lighting is crucial will blink the when taking selfies. Thus, we will use the led actuator to blink the lamp three times quietly suggesting a break. The chosen color will be cyan (the mixture of green and blue) - as it's a restful color.

- *It reasons that his is not urgent, so should not interfere with the person's activity.*

- Choose to beep the speaker gently at a low frequency indicating a reminder. Then, blink the LED in soothing green color slowly to communicate that the person needs rest. The wave arm and fan are excluded in this action sequence to avoid unnecessary distractions.

- The LED can draw the person's attention, blink at a calming slow pace and fade to a relaxing color. The fan can provide a gentle immediate physical feedback to remind the person to take a break, and the sound of the fan might also serve as a natural indicator and reminder of relaxation time.

- *The prompt does not state that the fan makes noise, it assumed that it has a sound.*