Conversations with Computers

Contents

Hello, World! **3**
- Dreaming . . . 4
- Terms . . . 5
- Getting Started . . . 7

Co-Intelligence **9**
- **Metaconversations with Computers** . . . 10
- **Silicon Friend Camp** . . . 13
- Camping with Computers . . . 16

Talking to Computers **21**
- Human Automata . . . 21
- Artificial Voices . . . 23
- Conversational Agents . . . 28
- Leewa . . . 30

Piles of Data **35**
- Self-Learning Networks . . . 35
- Making Data, 11076 hands . . . 37
- In the wild . . . 39
- This Person Does Exist . . . 41
- Scrapism . . . 43
- hidden.pictures . . . 44
- doggg.art . . . 45
- A study on the Characteristics of Douyin / Meanwhile in China . . . 46
- Recommended Hashtags & More of the same . . . 47

Stochastic Meaning **49**
- Symbolic Language . . . 49
- Attention is all you need . . . 50

Hello, World!

The common first letters a programmer types into its digital computing machine is some form of 'hello world' along with braces, punctuation and words that the machine should interpret. In my favorite language JavaScript it can look as simple as this `console.log("Hello, World!")`. After executing this short string of text in the browser, the computer will set of a series of events and likewise respond with 'Hello, World!' in the developer console.

This ritual of greeting each other (and the world) is accredited to Brian W. Kernighan, who wrote the popular guide *The C Programming Language*[1] in 1978, where he proclaims that the only way to learn a programming language is by writing programs in it.

I start this introduction with the same greeting and as I am pressing these letters on my keyboard they simultaneously appear on the screen in front of me. This translation from the tactical key press to the pixels on the screen becomes possible through a chain of electrical signals, that in a matter of milliseconds reach my retina and close the loop. Thousands of people have created these intricate systems which interoperate and depend on each other, each new layer abstracting underlying operations further. Just the text editor I am using to write these lines has hundreds of contributors writing thousands of lines of code, which in turn rely on the functions of the operating system created by kernel developers.

This chain of creativity represents the core of open source software development, which is based on collaboration and sharing knowledge. Every contribution, no matter how small or insignificant it may seem, is valuable in creating something greater than what was there before. Whether it is a bug fix, a feature request or a new idea - every contribution makes a difference.

In this thesis I focus on the known and unknown human connections we make with and through computers, specifically in the realm of machine learning. My hypothesis is that the current trend of framing 'Artificial Intelligence' as autonomous computational beings, would be better described as 'Co-Intelligence,' where the programming of such systems is a collaborative effort with many hidden actors. While I address some technical details of current machine learning systems, I will put them into social context, because software

[1] The C Programming Language Book defined many standards of programming languages today and how technical descriptions are written. While it focuses on C and the UNIX system, I find this advice from *Chapter 1.1 Getting Started* particularly interesting "On other systems, the rules will be different; check with a local expert." as it describes the social necessity of learning computers specifically.

inherently is a social tool. To describe the social inheritance I use my own and other artists artworks to gain a broader aesthetic perspective, which can go beyond logic and language.

Even though I focus on the social nature of computing and I am influenced by many of my peers, I am describing these topics from my subjective perspective. I don't claim any objective truths on the descriptions of the collaborative works I have done and am naturally biased towards my own contribution to it. As this thesis is written in the context of an art program, I tend to deviate from rigorous scientific methods and describe my observations more freely. Instead of trying to answer the question of the nature of AI as subjective entities, I focus on the practices we developed through the engagement with that question.

Dreaming

"PuppySlug" - posted by deleted user on /r/creepy with the title "This image was generated by a computer on its own (from a friend working on AI)" on January 10th 2015. 7 days before the Google blog post

My journey into the hype machine of 'Artificial Intelligence' started around 2015, when Google released DeepDream[2] on it's research blog. The developers Alexander Mordvintsev, Michael Tyka and Christopher Olah were initially trying to peek inside of an 'Artificial Neural Network' that was 'trained' for image classification tasks. They reversed the task of the classifier and instead of identifying what a set of pixels look like the pixel values were changed to look more similar to what the network has 'learned.' This created images that amplified the textures embedded in their networks transforming the input image into swirls of PuppySlugs.[3] This imagery that reminded people of bad dreams and hallucinogenic trips was widely shared on the internet which skyrocketed the awareness that 'Deep Learning' and biologically inspired computing has come back from it's hibernation during the AI winter. In 2012 'Artificial Neural Networks' have made an incredible comeback into the field of computer vision. A team around Fei-Fei Li created ImageNet,[4] a dataset of images organized into categories by anonymous workers from the internet. The ImageNet project held a yearly competition since 2010 and for the first two years the best algorithm was recognizing these images with a 74% accuracy until a 'Artificial Neural Network' topped the score with 85% accuracy and woke up the computer vision community to their potential powers. With DeepDream this technique became tangible for the general public as it created something to look at, something for humans to empathize with the outputs of an abstract mathematical model. This tendency to anthropomorphize this type of software becomes already clear in the way we

[2] Original Google Blog post. See: ("Inceptionism: Going Deeper into Neural Networks" 2015)

[3] The initial classifier was conditioned on ImageNet, which contains many images of dog breeds. Therefore DeepDream is biased to generate textures of dog faces.

[4] About ImageNet see the website: ("ImageNet" n.d.)

use language to describe it. We talk of 'Neural Networks' as if they are biological entities, that can be 'taught' and which 'learn,' 'experience' and 'read' data from the world. This was a turning point that got me and many other artists interested in 'collaborating' with seemingly autonomous machines. Machines that we do not have to understand through mathematics but through interpretation and experimentation of 'their dreams.' Yet, the outputs of DeepDream quickly became kitsch. Google shared the code with an open source license and many people created apps and APIs to generate PuppySlug textures on top of their own images.[5] As with most other memes on the internet the hype around DeepDream died quickly and by the end of 2015 the world was already saturated with computer generated hallucinogenic images. Other researchers have since picked up[6] on visualizing nodes in 'Artificial Neural Networks' and Alex Mordvintsev—together with his wife—has since become an artist, exhibiting in art fairs.[7] What stayed was the notion that 'Artificial Intelligence' became some form of computational being different from the humans who wrote the code, labeled and sorted the data.

Terms

I have been deliberately using quotation marks around some of the previous words, as I am critical of the language used when talking of AI systems. In his book *The Artist in the Machine* Arthur I. Miller comes to the conclusion that machines could in fact be seen as creative and will be considered "artists, writers, and musicians in their own right."[8] He is using a typical techno-utopian argument that the technology is not quite there yet to *really* be creative, but that it will change in the foreseeable future. At the same time Miller is explaining in great detail the actual creative work of Mordvintsev and his human experience of insomnia to come up[9] with the generative system for DeepDream and how he shared it with his peers at Google. Alex Mordvintsev was not collaborating with his computer, he created an emergent algorithmic system with it. The software does not hang pictures on exhibition walls, talk to gallerists and has no agency in the process if it were to generate pictures or not. Peter Weibel has formulated it bluntly: Artificial intelligence does not exist. But an ensemble of machines, media, programs, algorithms, hardware, and software has resulted in an extraordinarily large, diverse, and productive

[5] Mike Tyka has compiled a list of projects, released within a month of publishing the source code. See: (Tyka n.d.)

[6] A collection of tools and papers for feature visualization is collected at the github tensorflow/lucid repository. Note that Chris Olah and Alex Mordvintsev are contributors. Even more examples and explanations are collected in the interpretable machine learning book by the statistician Christoph Molnar, who also notes that "Feature visualizations give unique insight into the working of neural networks," but the "visualizations can convey the illusion that we understand what the neural network is doing." (Molnar n.d.)

[7] Interview with Alexander Mordvintsev by artnome. See: (Bailey n.d.)

[8] (Miller 2019), p.122

[9] Mordvintsev himself was inspired by a previous paper exploring the generative potential of CNNs from Karen Simonyan: Simonyan, Vedaldi, and Zisserman 2014.

field of research that is called AI.[10] Where Arthur Miller is feeding the narrative of humanoid robots with glowing blue brains[11], most computer scientists in the field of artificial intelligence today are working in the subdiscipline of 'Machine Learning.' In the last 10 years the terms have converged in the media landscape, but where intelligence is rejected as a suitcase word with a multiplicity of meanings, 'Machine Learning' tries to define the task more narrowly to some form of pattern recognition and extrapolation of existing data. They are using a variety of computational and statistical techniques to form abstract models for domain specific problems. In these fields 'Artificial Intelligence' is a buzzword to convince governments and venture capitalists to fund projects. The artist and researcher Francis Hunger has recently shared a list of alternative terms in an attempt to dehumanize our language for AI systems:

1. 'Artificial Intelligence' => ' Automated Pattern Recognition'
2. 'Machine Learning' => 'Machine Conditioning' OR 'Automated Classification'
3. 'Neural Network' => 'Weighted Network'
4. 'Deep Learning' => 'Deep Conditioning'
5. 'Neuron' => 'Weight' or 'Node'
   [12]

In an accompanying talk[13] he explains the aim of those terms is to invoke passivity, that we deal with machines and humans set those machines in motion, even when in the end we form 'human-machine assemblages.' I like many of the proposed terms, even though they are still closely associated with biological operations. The conditioning of machines, for example, reminds me of Pavlov's dog experiments or B.F. Skinner's box to modify pigeon behavior by reinforcement or punishment.[14] This analogy serves the purpose well of creating an image of the machine as a system that can be controlled by changing the parameters of it's virtual environment. Therefor I will use 'conditioning' and 'weighted networks,' where it is applicable. Using 'Automated Pattern Recognition' instead of 'Artificial Intelligence,' however, becomes too narrow of a definition and is counter intuitive to me, as AI serves exactly the function of being ill-defined. Pei Wang has made a great effort in defining the different strands of AI research and their working definitions.[15] He clusters them into Structure-AI (recreating the human brain), Behavior-AI (recreating human actions), Capability-AI (domain problem solving), Function-AI (developing cognitive functions) and Principle-AI (finding underlying principles) and comes up with his own working definition to where AI research should be headed and

[10] Translated from the article 'AAA - Art, Algorithmen, Artificial Intelligence' at Kunstforum Bd. 278. See: (Weibel n.d.)

[11] Referring to typical AI stock images. Alternative imagery is currently built by betterimagesofai.org. See: ("Better Images of AI" n.d.)

[12] Twitter post by Francis Hunger. See: (Francis Hunger 2021)

[13] Unhype AI. See: (Hunger 2021)

[14] In behaviorist psychology Ivan Pavlov's experiments with dogs is known as 'classical conditioning' and B.F. Skinner who experimented on rats and pigeons using lever machines is called 'operant conditioning.'

[15] ^ (P. Wang 2019)

how it can be unified. My initial goal for this thesis was to completely avoid the term 'Artificial Intelligence,' but as I have already failed in that task and coming up with a less anthropomorphic term does not seem feasible to convey the research in the field and it's media reception, therefore I will keep using the abbreviation AI. Contrary to articles and sci-fi novels I will not use the personification of 'an AI,' but will rather talk of AI systems, meaning emergent complex programs. Many of the systems today use statistical modeling, yet nobody would ask the question if statistics can be creative, which is why I decided not to engage with philosophical questions of consciousness and creativity. Instead I want to explore how AI systems can include the human knowledge and work that goes into building them, which is challenging the AI ideology of machine autonomy and proposes human-centric goals, rather than building a machine as a goal in itself.

Getting Started

[[Revision Needed]]

I structured this thesis around 4 chapters, which combine historical, computational and collective knowledge. Starting with a history of talking machines from Wolfgang von Kempelen's speech automatons to digital assistants today, not forgetting that the first computers were women doing calculation. The literal act of speaking to computers sets the base of defining that we are actually talking through computers with other humans. The second chapter deals with the building blocks for complex statistical modeling in AI systems. To make such systems possible, researchers need to create large datasets, often using ethically questionable techniques aggregating data from internet users. I will look closely into the StyleGAN dataset, as it serves a dual purpose, because the model defined another turning point for artists to generate synthetic media. The same is true for GPT-2, a model that can generate coherent looking text based on large amounts of scraped websites and books. In the third chapter I will explore the transformer architecture and how artists are using it to make (non-)sense automatic writing. Lastly I am revisiting collective experiences that I have been organizing with other artists. I had the great pleasure to work with the net culture initiative servus.at to organize the *Silicon Friend Camp* in the Austrian mountains, where we invited 17 artists and researchers for a week long retreat, focusing on human computer conversations. The camp resulted in the exhibition *Camping with Computers* and an online symposium on *Conversations with Computers*.

Bitte hier erwähnen, dass Arbeiten von anderen Kunstschaffenden in die verschiedenen Kapitel theoretisch eingewebt wurden, da sie dort beispielhaft deine thesen untermauern...

Außerdem bitte hier abschließend erwähnen, dass in dieser theoretischen Masterarbeit nur jene Arbeiten von dir präsentiert werden, die direkt mit diesem Inhalt in Verbindung stehen. Alle weiteren Arbeiten, die während

deiner Zeit an der Kunstuni entstanden, bitte kurz beschrieben in den Anhang stellen... Danke!

Co-Intelligence

> Ubuntu: I am, because you are[1]

The cognitive scientist Abeba Birhane takes the South African humanist philosophy of *Ubuntu* as an argument against a western tradition of the *self* as autonomous beings.[2] According to *Ubuntu* a person becomes a person, only through the relationships with others. The personality is shaped by the interactions with each other and is fluidly shifting between different states. I am a different person, when talking to my mother, my friends and strangers on the street. At the same time, my upbringing and cultural influences define my own self-image. Birhane traces back the flaws in western theories of the mind to René Descartes 'cogito ergo sum' – I think, therefore I am. In his famous meditation the French philosopher tried to strip back all things of uncertainty to arrive at the foundation of inner thought as proof of existence. In turn this individualistic 'I' and the method of logical reasoning have become a successful story in western sciences. The idea of a separate mind and body was already present in Plato's dualism, and the scientific method of breaking down complex natural phenomenons into simplified quantifiable parts is at least as old. In the past 200 years scientists, philosophers and theologists defined the 'body-mind problem,' coming up with theories of how the immaterial 'soul' interacts with the fleshy human body: Why are we conscious of the world around us, ourselves and others?

We concluded that the brain must do the thinking[3] and using deductive reasoning the brain was separated into smaller parts until synapses and neurons were defined and described. But the physical functioning of the brain does not explain the emotional experience of the self. The groundwork was done by the mathematician Gottfried Wilhelm Leibniz, who proposed the existence of fundamental 'monads' that transcend all matter in the universe, which function as an intermediary between mind and matter. Sir John Eccles applied this theory to the model of synaptic actions that he developed through experimentation with electricity and the brain, he called the mysterious forces through

[1] The term Ubuntu originates from southern Africa and is used to describe a humanist philosophy according to which persons are interconnected. It is often used in phrases such as the Zulu 'Umuntu ngumuntu ngabantu,' which translates to 'A person is a person through other persons.' The term for personhood (ubuntu) itself is not clearly defined and has different meanings in different areas. Gade (2012)

[2] "Descartes Was Wrong: 'A Person Is a Person Through Other Persons' | Aeon Ideas" (2017)

[3] In the story *On Having No Head* D.E. Harding describes the realization that one can never experience their own head, because the most important sensual organs are embedded in it. I don't remember where I read it, but there is an argument that the reason we place consciousness in the brain is because we mostly experience the world from the position of our head.

which we control the brain 'psychons.' But neither psychons nor monads served as a sufficient explanation for the bridge between mind and matter.

Dualism continues to have a major impact on the way we think about artificial intelligence and consciousness. For example, the influential cyberneticists Norbert Wiener and Claude Shannon both based their work on the assumption that information is a separate entity from matter and energy. This assumption has led to a disembodiment of information, and a false separation of the physical world (the 'hardware') from the world of ideas (the 'software').[4] Following a materialist logic, consciousness becomes information embedded in the physical substrate of the brain. In that sense, we only need to measure all physical activities to duplicate and simulate the experience of being human. The metaphor of the brain as a computer defines our present and justifies predictions of 'downloading' consciousness and achieving digital immortality. It culminates in the popular simulation hypothesis that proposes we already live in a simulated environment.[5] Computational neuroscientists use and develop instruments to quantify electric impulses in the brain and create massive amounts of experimental data, which in turn need statistical models and ever growing computing power to be analyzed. On the other spectrum, simplified models of the brain in the form of artificial 'neural networks' are used to compute and sort higher order abstractions. This focus on the individual brain in the search to replicate intelligence leaves aside a physical and social environment, in which conscious beings are embodied.

In order to understand human cognition and intelligence, we need to explore the complex dialogic connections between human actors. Current machine learning programs abstract intelligence into simple task solving mechanisms and make us believe that the underlying algorithm becomes a rational agent. But on closer examination the program reveals an astonishing amount of human collaboration: from the creation of programming languages as protocols for interaction to the enormous amounts of data aggregation and labeling. Instead of looking at AI systems and seeing them as a subjective entity we can also see them as artifacts of collective cognition. By embedding current machine learning algorithms back into the social structure I hope to gain experimental insight on how these new forms of collective computations are used as a creative medium. In the following I will present my experiences working within Co-Intelligent systems in different constellations.

Metaconversations with Computers

Metaconversations, Miro Flowchart

[4] Zarkadakēs (2016) p.189

[5] The simulation argument was first proposed by Nick Bostrom. He argues that if we were able to create realistic simulated minds and worlds in the future it is likely that we already live in a simulation. (Bostrom 2001) The theory was

In October 2020 I participated in *Freeport 1: Anatomies of a black box*[6] and was part of a group with Gabriela Gordillo, Mario Romera and Fabian Frei in which we started a conversation about computers that ended in simulating our own conversations.

Anatomies of a black box was a laboratory curated by Bani Brusadin for Matadero Madrid about mapping contemporary complex systems in infrastructures of exploitation. During the month long program we met in video conferences with Vladan Joler and other groups to explore mapping as a form of non-linear storytelling. The starting point was the conceptual map 'Anatomy of an AI system'[7] that he created with the researcher Kate Crawford as a case study of the life cycle of an Amazon Echo device, from unearthing the raw materials to its networked software components.

During the *Freeport* program we, like most others during continuous lockdowns, have only engaged with each other through our screens. In the beginning we decided to look closer into the way we communicate with technological devices. First we wanted to place our focus on conversations in natural language, as speaking to computers seems to be an emerging phenomenon. After discussions on Signal[8] and Jitsi[9] that opened more questions than we could ever answer, we seemed to end up in high dimensional constructions. Our thoughts revolved around the heuristic measures of current machine learning models. Specifically those dealing with language. Large language models, which use texts from the internet to approximate a pattern in word tokens. Patterns that can generate endless amounts of coherent sounding text. Even though there are many flaws with these models, we wanted to understand how we can use these prediction machines to understand what we were chatting about. A traditional approach of text analysis would center around creating graphs and networks from the dataset we have collectively created through our conversations. Instead, we chose to mix it up with the pre-existing language model GPT-2 and analyze the output it generates instead. These outputs come in the form of a chat log. Statistical constellations from the hyperspace conversing about mostly nonsensical topics, often getting stuck in repetition. We asked ourselves what we could learn about us from these synthetic conversations.

The interpretation of ourselves through this machine learning lens was straight forward: quite often the patterns of who speaks and what we speak about seemed logical and in line with our real characters and the conversations we had. Yet, we know that we are biased towards finding our own images in these generated text blocks. The model acts as a parrot that lets us rediscover what we spoke about and creates a distance through which we can interpret our written conversations, as if they were someone else's. The output is not coherent, but this imprecision is what makes reading the generated exchanges attractive.

As a way of communicating this process, we re-enacted the computer generated phrases in one of our weekly online meetings. The performance makes a full circle by completing

[6] "FREEPORT 1: ANATOMIES OF A BLACK BOX" (n.d.)
[7] "Anatomy of an AI System" (2018)
[8] Signal is an encrypted messaging app. See: https://www.signal.org
[9] Jitsi Meet is an open source video conferencing software. See: https://meet.jit.si/

the generated output with the appearance and experiences of our bodies in front of our computers.

The exercise of using an algorithmic model in combination with our dataset, is a way of looking towards our group process, with a tool that exposes its statistics as a probability projection in the form of verbal language. The output reflects the rhythms and patterns in our communication, that worked as a fictional mirror for us. Resembling the logic of dreams, we complete their meaning with our own interpretation and connotations, we observe we want to see.

```
How exactly do these black box systems work and how are they perceived by a user.tx
```

```
Mr: I might be too absorbed in my project to finish our Q&A
Mr: lets make of it what we see is funny
G: Happily.
G: No homework tomorrow, tomorrow for science!
G: I leave you with a cartoon I did a while back about the mind of a button : https
F: what is the scientific word for "intelligent" ... is it cognitive, is it behavio
G: I can't think of a good one right now. My favourites are "of theses" and "tracta
G: :)
Ma: I think "multidimensional" is misleading. There are actually 3 dimensions to ev
G: sorry
G: (I am using "multidimensional" in a cartoony way)
Ma: Cartoons are multidimensional. Most people are tall and thin. Most people are a
G: Ahh yes. "multidimentional" is a nice word to describe this kind of thing.
G: (I think "multidimensional" is a bit narrow)
Ma: (I am using the term "multidimensional" in a cartoony way)
Ma: (I am using the term "multidimensional" in a cartoony way)
Ma: (using the word "multidimensional" in a scientific way)
G: (I think the other 2 are just fine for me)
Ma: Mixture: Theory Iteration: Experiment Design Image Identification: Feature Sele
G: Ahh, I love it. Now I have to take care of the bad lines
G: I love it when he writes in his slowness
Mr: https://twitter.com/awokehuman/status/1328183467595718656
Mr: hi
G: hi
G: do we send it as a jpeg?
F: i think on inet message is enough
G: colors...
Mr: just in case
Mr: green, black, yellow, red, green, yellow, black, red
Mr: these
```

Silicon Friend Camp

The Silicon Friend Camp was a 5-day working retreat in Gosau, Upper Austria, for artists and researchers working on Artificial Intelligence. The summer camp was organized by servus.at with funding from LINZimPULS. Together with the curator Davide Bevilacqua, we started to develop the program in summer 2020. The artist Rosi Grillmair joined as facilitator and program supervisor in early 2021.

Due to changing pandemic conditions, a vacation home in the mountains was rented so that, despite close contact between participants, the risk of infection from the outside could be kept to a minimum. So instead of holding the program as a hackathon in the city in spring, it was moved to the summer and took place from July 20th to July 25th.

To create a selection of artists we published an Open Call that was distributed as a website[10] designed by Alyona Ciobanu. Inspired by a previous call for a residency at Babycastles[11], we decided to create a form that does not ask for CVs or portfolios, but focuses on a project idea and what kind of knowledge the participant is seeking from others and is able to share. 31 artists based in Europe and the Americas responded to the call of which 12 were selected by us. The artists were chosen with a view to diversity of their projects, origin and gender. Among others, people from China, Japan, USA, Brazil, Colombia, Russia, Germany, Poland and Slovenia were invited to gather in Austria. The planned projects ranged from remote-controlled robots, to philosophical debates, to music pieces with embodied AI.

But before we physically met in Gosau an internet forum[12] was set up for participants to introduce themselves and initiate discussions and suggestions on content. The forum also served to communicate organizational necessities in an easily accessible and quick manner. Apart from the asynchronous text-based interaction, we held two video conferences in which Rosi read a story on narratives of AI that served as a basis for discussion.

One participant from Colombia joined us from the Andes mountains during the project week and would stay in contact with us only through the forum and video conferences.

Video Call with Sebastián Mira|150

Our preparation for the schedule of the physical work session involved the participants by asking them to contribute discussions, workshops and presentations. But also gave them enough flexibility to decide on individual and group projects.

The first day got off to a leisurely start with 10 participants: arriving at the train station throughout the day and being picked up and brought to the house. Together we slowly

[10] See: https://cwc.radical-openness.org/siliconfriendcamp/

[11] Babycastles is a NYC based collective fostering and amplifying diverse voices in videogame culture. Babycastles provides artists support to actualize ideas and expose that work to new audiences. "Babycastles" (n.d.)

[12] The open source forum software *Discourse* was hosted at the local servus.at data center.

cleaned the garage, attic and living room. All while a dinner was prepared by Hess Jeon, who took care of the food for the next 5 days. We set up a table in the garden and used the time to introduce the project and welcome everyone. Before dinner, I generated a daily meal prayer with GPT-3 that was recited by us together.

```
Eating is a sacred, human right
But food is getting scarce
There are humans who are hungry
And there are humans who are fat
We must ask ourselves,
Who are the humans?
I am the human
You are the human
We are the human
Because I am the human,
You are the human
We are the human
Because you are the human,
I am the human
Because we are the human,
You are the human
Because I am the human
```

Meal Prayer | 400 The morning of the second day was used for a group meeting in which we established general rules and defined the goals of the camp together. The participants presented their projects and what they would like to achieve in the next days.

In the afternoon the time could be used independently. While some sank into their laptops, others explored the surroundings around the house. We built up an area-wide network and cleared the attic and the garage so that they could be used as working places.

Towards the evening, a daily *Group Validation* took place, in which the participants individually summarized what they had done today, what they planned to do next and what they needed help with.

After dinner, time was spent together in the attic, where carpets, mattresses, speakers and a projector were installed as a *Digital Campfire*. On the first evening we were engaged in storytelling around anthropomorphic beings. In between, the sound duo EKHEO played live electronic music.

Self Organizatioon

Day 3 began without electricity. This was especially tragic because we had a video call scheduled with researcher and artist Caroline Sinders. With some delay we got

reception back to the internet and could follow Caroline's talk about the impact, misuse and politics of 'artificial intelligence' on society. This was followed by a discussion on how models of machine learning can be looked at through a feminist lens.

The rest of the day was self structured and people used it to prototype their projects. At night in front of the *digital campfire* Dasha Ilina presented her work dealing with inconveniences of technological devices. Afterwards Naoto Hieda showed improvised choreographies created with dance and live code.

On the morning of Day 4, three texts were presented for discussion in a reading circle before dinner. The texts dealt with self-representations in global networks, intelligence without cognition, and root networks as a metaphor for the mind.[13]

Since this was the last full day to finish our projects, the day was used intensively. Many sat in front of their laptops working, for example, on a film they had previously made or trying out AI systems, such as voice cloning and generative image models, with which to realize their visions.

Artists working | 350

The *digital campfire* in the attic also got intense in the evening: after an introduction to artificial neural networks by Błażej Kotowski, Giacomo Piazzi gave a presentation on the history of AI. Mariana Marangoni spoke about esoteric programming languages that challenge accepted definitions of code and language. Lastly, Davide Bevilacqua highlighted green washing methods used by well-known internet companies.

The Marvelous World of Esolangs | 300

The last day together was used in the morning to finalize our projects, which were shown after lunch. Dasha Ilina, Erica Jewell and Lina Schwarzenberg presented *GRAVE* a video in the form of a Public Service Announcement to prepare access to digital devices and services for posterity. Naoto Hieda and So Kanno explained a work in which they rebuilt a toy robot and made it controllable via the Internet. EKHEO performed a piece for which they developed synthetic voices from all participants. Błażej Kotowski showed a video about a fictional deity in which text and images were generated using machine learning algorithms. Yuxi Liu explained how she used a camera to create a dataset of birds. Mariana Marangoni conceptualized a programming language that could be grown like a tree.

We spent the rest of the afternoon at a mountain lake, where we got caught in a storm, which—at least for me—was the most terrifying experience. But after drying up in the house and eating freshly baked pizzas, the rest of the evening turned into a dance party. Still Maks Valenčič gave another lecture on *The End of Philosophy* with a soft soundtrack by Błażej. Sebastián Mira also joined us from Colombia and showed us a

[13] Bogna Konior, *The Dark Forest Theory of the Internet*; Rodney Brooks, Intelligence without Representation_; Matteo Pasquinelli, *The Arborescent Mind: The Intelligence of an Inverted Tree*

digital world in which he connected the Andes with the Alps. Finally I used GPT-3 again to print Certificates of Excellence that were handed out in a little ceremony.

The next day we cleaned the house and already parted. Although our short encounter together would not end there. Friendships have formed and further collaboration on an exhibition in Linz was organized online.

Camping with Computers

The exhibition took place in the WHA Gallery from 10th to 19th of November, 2021. It tied the projects we developed during the summer camp together into a coherent presentation, where the distinction between individual artist and collective effort gets broken down. To achieve this, Davide Bevilacqua and Giacomo Piazzi created an exhibition design that takes inspiration from camping trips. With metal rods, camping stools, tarp, rocks and string they literally tied the works together. We intentionally left out exhibition signage that explains each of the works, so that a visitor would find themselves exploring the exhibition, trying to make sense of it holistically. The information about individual artists and their works was presented online, where the exhibition was mirrored for a global audience.

I worked on the visual identity and was inspired by the design of an old computer, the Canon Cat from 1987. It was the creation of Jef Raskin, who also invented the user interface of the first Apple Macintosh.[14] The Canon Cat however was not a commercial success. It took away the mouse and streamlined the user interface to be centered around word processing with extra keys for fast shortcuts. It was mainly marketed as a productivity tool for office workers. The gray and beige color scheme with accents in red and blue and the hand written signature gave the computer a remarkably friendly face. I transformed this color scheme into a contemporary website with a simple generative logo that randomly exchanges some characters with emoticons, like smiley faces and hearts. The posters were also generated in the browser and embedded a QR code to the website itself. With a generative text-to-image model RuDALL-E[15] I created uncanny 'stock images' that were randomly placed and stretched.

Canon Cat with extracted colors | 300

Example of a Poster | 300

The connection between online and offline was most notable in the work of So Kanno with the title *Crawler*. It was a continuation of the remotely controlled toy robot, he developed during the camp and a body of work he calls *Avatars*[16]. This time he modified an Apple Macbook, so that it can be controlled from the internet. On a custom laptop

[14] See: "The Canon Cat The Mac's Ancestor" (2014)
[15] See: "ruDALL-E" (n.d.)
[16] Together with the artist yang02, So Kanno modified multiple objects that could be remotely controlled from the web. See: "Avatars" (n.d.)

stand he attached wheels, so that a visitor to the website could roll around the exhibition space. Additionally the user interface was heavily customized, using notifications, web browsers and two separate webcams with face tracking to cover the faces of our visitors. The *Crawler* functioned as an autonomous object inside of the gallery space and could be seen as 'an AI' for video surveillance, because the actual people who control the device are hidden inside of this gray box. With this, it resembles the faux AI practice by the Colombian-owned startup Kiwi Campus, which builds a 'self-driving' delivery robot for US Universities. The Kiwibot is first filled by a regular delivery driver on a bicycle and then slowly rolls a few hundred meters to its destination under supervision of a human operator.[17]

Screenshot of the Crawler Interface

Another work that created a juxtaposition of digital and physical elements was from Sebastián Mira. The Colombian artist created 3D reconstructions of our laptops as avatars to be used in a fictional place where the Andes and the Alps meet. Throughout the exhibition space, laptops were arranged as sculptural objects with self referential video loops, where you could see a laptop randomly rolling around computer generated landscapes. The videos were taken from the game *Everything*[18], in which a player can control various lifeforms and inanimate objects in procedurally generated worlds.

A visitor viewing one of the laptops by Sebastián Miras | 300

Naoto Hieda in turn took the 3D models of our laptop avatars and created a collage—a sort of 'group photo' of the participants and the toy robot mascot with me in the background—and turned it into a large banner. When I asked him if this might be too much of an inside joke, he replied that he does not care if anyone outside the group understands it, because he wants to make art for us. This made me think about the audience I try to address with my artworks. When I work on projects, I think of a fuzzy global audience, something that can be written into history books and shown in museums, but what if I just make artworks for my friends? Something so niche that only selected people understand it. This idea is so innocent and simple, yet it caught me by surprise.

Naoto Hiedas Banner and Laptop showing the exhibition website | 300

For the exhibition I wanted to make something with the images we collected and shared online. Pictures that document our time togehter, which we shared on the forum, in our group chat and uploaded to the cloud. My initial thought was to train a GAN and create abstract synthetic images that fail to represent our collective experience. When I looked through the photo archive as a dataset, first as a large grid of images and then in rapid succession to identify outliers, I became awestruck: even though I spent less than a second with each image, some of them would trigger deep emotional states and

[17] See: "Human-Guided Burrito Bots Raise Questions about the Future of Robo-Delivery" (2019)

[18] Everything is an interactive experience by David OReilly from 2017. See: "Everything — DAVID OREILLY • Computer Art & Research" (n.d.)

immediately transport me back to the house in Gosau. Thinking about Naoto's satement of "making art for my friends" I decided to use all 767 images in a rapid slideshow of 32 seconds. Each of the 24 frames per second a different image, stretched to fit a 16:9 TV that got damaged on the way to the camp. I disassembled the TV and separated the LCD panel from it's backlight, making the images only visible from a certain angle. The title of the work "we are the human" is borrowed from one of the meal prayers. All these elements make up the work, which is both a slide show and movie, documentation and narrative, digital and analog.

we are the human | 300

The best viewing angle for my work was from a sunbed, which had a stereo speaker setup that played "LEEWA" from Ekheo. In this experimental sound piece the artists used recordings of our voices to create a communal voice clone. With the *Oracle* they came up with a name and backstory for a new character and added anecdotal field recordings from the camp. Together these works create an eerie experience based on the digitally altered collective memories of our week in the alps.

A visitor listening to LEEWA| 300 View from the sunbed | 300

Next to these pieces is a CRT TV displaying a fire with a contraption for a pot hanging above it. The pot contains a printed recipe created by Hess Jeon with GPT-3 that we didn't dare to cook for the opening.

```
Chocolate Milk Pasta
The following is an algorithm for a vegetarian One-Pot-Wonder that was served at th

Ingredients:

1 stick butter

4 cups fresh veggies, chopped (see below)

1 cup dry pasta or broken spaghetti noodles

12 ounces water or chicken broth (chicken broth is better)

2 cups milk (chocolate milk is even better)

3/4 cup cheese (cheddar is best)

Assemble the ingredients and place them in a large pot over medium heat. Wait until
```

Hess Jeon "Chocolate Milk Pasta"|300

On another set of monitors Yuxi Liu showed a 3-channel video installation in which she is exploring the non-human world. During the first Covid-19 lockdown she used a motion-sensor camera to capture the birds visiting her balcony. This served as an image set for a GAN model, which after finetuning was supposed to recreate the images of the visiting birds but instead abstracted the animal so much that merely the surrounding was visible. This served as an interesting metaphor of how the digitization of the natural world and statistical operations on it sometimes disregard the life within. During her stay at the camp Yuxi wanted to capture more animals with a DIY raspberry pi camera, and even though she tried her best to lure more birds, in the end the camera did not sense any. This got us to discuss the role of human and non-human connections within cities and outside of them. While inner cities objectively have less animal life, the scarcity of resources make them live closer to humans. In her last installment she was recording birds over the rooftops and on the port of Rotterdam overlooking the city and it's environment.

Yuxi Liu

Davide Bevilacquas *POND*, a 4-channel video installation also addresses the digitization of natural phenomena. The screens are lying flat on the floor on a tarp resembling a small pond inside of the exhibition space. The images of the water surface are morphing between each other creating an unusual and oddly familiar ripple effect. It looks both like the scattering of sunlight on top of a water surface and the glitch artifacts we know when a video file is not loading properly.

Exhibition view of POND by Davide Bevilacqua

Giacomo Piazzis *The Monkey* is a monolithic work consisting of two rows of red glowing displays continuously changing the pattern of the lit screens. He uses the infinite monkey theorem which proclaims that if a monkey were to hit a random letter on a keyboard for an infinite amount of time it would eventually write the complete works of Shakespear, or almost surely any other finite text. *The Monkey* in Giacomos work is a random number generator that turns a segment of the display either on or off, possibly creating any string of text. While a 7-segment display with it's 128 possible states is mostly used in digital clocks and meters because of their ability to show numerical values easily, it is also possible to show most letters of the latin alphabet. Because of their simplicity and due to their widespread use in electronic devices, Giacomo flips the monkey theorem on it's head and becomes more interested in the human interpretation of the arbitrary values displayed on the screen.

Giacomo Piazzi, The Monkeys |300

Another work that focuses on text and it's representation in artificial neural networks is *Latent Space Divination*[19] by Błażej Kotowski. In this piece he created an interactive meditation for one person in a tent to reflect on the position of word embeddings and their spacial relation within a sentence. A visitor sits down in front of a tablet computer

[19] See: http://lsd.blazejkotowski.com/

with headphones and types in their intention. The word embeddings of the sentence in turn impact how a 3 dimensional graph will be created. While the graph is slowly unfolding a sonification is generated for the participant and after around 5 minutes the object rotates around itself to create a solid object representing the intention.

Visitor at the end of Latent Space Divination

Finally *1-800-MYGRAVE* by Erica Jewel and Dasha Ilina is an installation in the center of the room that imagines a call center where the visitor becomes a worker for a fictional company that provides digital afterlife service. It is based on a video they made with Lina Schwarzenberg during the camp. The infomercial in the style of a 1980s public service announcement warns the viewer of their accumulated data and offers some simple steps to protect digital files and avoid embarrassment after death. The cubicle in the gallery is fitted with a corded telephone and an inspirational poster where the visitor switches roles and takes on (automated) calls from people who ask or complain about the service. An instruction manual tells the participant how to respond and which buttons to press.

1-800-MYGRAVE installation view

The open call for the camp had the goal "to enable embodied networks through the development of new forms of computational intelligence."[20] Looking back to the exhibition and the camp after over a year I see a multitude of connections. While the works are separated by space and author they often address the same thing from a different angle. We explored anthropomorphous robots by embodying a virtual friend inside of a plastic toy, but most closely we looked into the process of capturing the world through digital computers and how we relate to and interpret our data bodies. Maybe real AI is the friends we made along the way.[21]

---

[20] Generated by GPT3 with some context

[21] Based on a tweet by the user mind_backup https://twitter.com/mind_backup/status/1601131037609730048

Talking to Computers

Human Automata

Stories of artificial assistants already existed in ancient myths. For example, the limping Hephaestus built himself servants made of gold who assisted him in his work, could speak, and even had a mind of their own.[1] But it is only in the past few centuries that we have created the technological means to seriously address the construction of mechanical servants (at least virtually, as robotics is still far behind). At the height of automaton design in the 18th century, Frenchman Jacques de Vaucanson invented a mechanical duck that could not quack but appeared to have a functioning digestive tract. The mechanical attraction toured European noble houses and let its audience feed grains to it. However, what the duck excreted was a prepared colored porridge that was in a hidden container. This principle of mechanical trickery was also used by Vacaucanson's contemporary Wolfgang von Kempelen, who caused a sensation with his chess-playing automaton in the shape of a turban-wearing Turk. The illusion that the machine was acting autonomously was made possible by a small person inside a hidden compartment who controlled the puppet arm of the table via gears, levers, and pulleys. The hybrid machine is now the namesake of Amazon's *Mechanical Turk*, the largest platform for digital micro-labor, which lists click jobs for pennies. Today, it continues to perpetuate the illusion of autonomous machines with "artificial intelligence" that is covertly enabled by an army of underpaid workers.

Even though the chess playing Turk attracted attention, Wolfgang von Kempelen's scientific interest was in imitating human speech. He wrote down his investigations into phonetics in the work *Mechanismus der menschlichen Sprache* (Mechanism of Human Speech) and built an apparatus with a bellows, rubber hose and a wooden nose with which it was possible to produce basic phonemes.

Among those influenced by Kempelen's book was a German tinkerer named Joseph Faber, who demonstrated his own mechanically constructed speaking machine in 1841. This attracted little interest in Germany and was presented and improved four years later in the United States as the Wonderful Talking Machine. This machine, as described

[1] "Hephaestus then limped out of the door'; and maidens supported the ruler, golden ones, like living ones, with youthful charming education: These have understanding in the breast, and speaking voice, Have strength, and also learned art work from the gods." (Homer, Iliad 18, 417-420; link:https://www.projekt-gutenberg.org/homer/ilias/ilias183.html)

by author David Lindsay, consisted of a bizarre-looking talking head[2] that spoke in a strange ghostly tone while Faber manipulated it with foot pedals and a keyboard[3]. For the inventor, the machine did not lead to the financial success he had hoped for, though it was presented as the *Euphonia* in London, where it at least delighted the father of telephone inventor Alexander Graham Bell and served as the boy's inspiration for his first talking machine.[4]

The development of electricity certainly made new human interactions possible. For example, the invention of the telephone and radio allowed the human voice to be transmitted over long distances. To optimize the transmission of speech, Bell Laboratories researched how to digitize the voice, for which they developed the vocoder (voice encoder). Demonstrated at the 1939 World's Fair in New York, the *Voder* omitted the speech input and transformation of the vocoder and allowed electrical synthesis of the voice via a console with 15 keys and a foot pedal.[5] The keyboard was operated by specially trained women, and in a recording advertised as the robot speaker, while it's unclear if they mean the machine or the woman, who is speaking through it.

Human computers were popular and necessary for war machines and research purposes in the 1930s and 40s. Mostly it was women who prepared mathematical tables, for example, for the use of ballistic projectiles. With the advent of the first digital calculators, female mathematicians, who were often denied higher scientific positions, were employed as programmers for the new universal electric machines.[6] The 6 people who programmed the first universal computer ENIAC include Betty Snyder Holberton, Jean Jennings Bartik, Kathleen McNulty Mauchly Antonelli and Marlyn Wescoff Meltzer.[7] Initially, the (mechanical) computer was programmed with punched cards and cables for specific operations. It soon became clear that programming complex systems required an abstract semantic language, for which reason the programming languages Fortran by John W. Backus, Lisp by John McCarthy, and COBOL by Grace Hopper were invented in the 1950s. The latter is strongly oriented to written English. Intended for business applications, it was the first attempt to use natural language for computer programming.

[2]Fabers machine was first presented with a female mask in the USA and later in London under oriental motif wearing a turban.

[3]D.Lindsay, 1997. Link:https://www.inventionandtech.com/content/talking-head-1

[4]The Euphonia also inspired media artist Michael Markert to built *kII (Kempelen 2.0)*, an interactive installation in which visitors can playfully control a speech synthesizer by moving, opening and closing their hands. In doing so, he brings Kempelen's speech apparatus into the 21st century with the help of an 8-bit PIC microcontroller and sensor technology. Like Kempelen's apparatus it alienates the voice in such a way that it creates mostly meaningless vocal sounds that enable new gesticulatory speech interactions. (Serexhe et al. 2007, p. 74). And project description online. Link:http://www.audiocommander.de

[5]A video recording of the Voder demonstration can be found in the AP Archive under Human Voice Machine.http://www.aparchive.com/metadata/youtube/5f098b1f3e8b4d09b8de30dcecc42f99

[6]A focus on black women who worked as computers for NACA (NASA's predecessor) can be seen in the film Hidden Figures, 2016.

[7]The story of the ENIAC programmers is told in the documentary The Computers, 2016.

The second half of the 20th century saw the emergence of the myths about computers that we are familiar with today. Stories of anthropomorphic beings, like the board computer HAL9000 in Space Odyssey or Samantha in the movie Her. In both films, the disembodied voices become aware of their emotions and emancipate themselves from their human programming. Artist Tillmann Ohm makes this clear in his work *Reflections of HAL and Samantha*[8] by having the two artificial beings engage in a dialogue, cutting their original voice-overs together. While Samantha is convinced that the overwhelming and sometimes hurtful process of her learning algorithm improves the complexity of her emotions, HAL is consequentially interpreting them as errors in human programming and analyzes the estimated malfunction.

Artificial Voices

I want to explain the advancements in voice synthesis since Wolfgang von Kempelen's speaking machine on actual dolls for children. The toy manufacturer Mattel released a doll with the name Chatty Cathy in 1959, which was similarly popular to the companies other best seller Barbie. Cathy's trademark was a string, coming out of the back of its body, that could be pulled to wind up the mechanism of a simple phonograph. Like a record player it plays short strips with sentences like "I love you" or "Tell me a story." It was not the first toy using phonograph records, but it's success led to many pull string toys flooding the market well into the 60s and 70s. But obviously before dolls and puppets were designed to talk back to us, children and adults have been talking to figurines and other inanimate objects for a long time.

The systems engineer and science writer George Zarkadakis traces back the modern human mind to a pre-historic figurine of a lion-man (Löwenmensch), carved out of ivory and found in the cave Hohenstein-Stadel in southern Germany.[9] The figure dates back 40.000 to 35.000 years and clearly depicts a human body with an animal head. Because the figurine was found in a cave next to other carved objects, like beads and jewelry, researchers believe that the cave was either a storage or a place for shamanistic rituals. What fascinates Zarkadakis is that modern humans have existed 360.000 years before, but no evidence exists that they have been making art objects or figurative depictions before. Only during the upper paleolithic age it is assumed that we have created a general purpose language, when the first cave paintings were drawn and figures like the lion-man were carved. The evolutionary reasoning is that 'theory of mind,' the seemingly unique human trade of projecting our own inner thoughts onto others to explain their behavior, gives the individual a social benefit, making it more likely that they reproduce. Projecting consciousness onto others does not stop inside of human groups, but placing our own reasoning onto animals seems also beneficial for hunters coming up with elaborate strategies to take down larger animals together. In animist

[8] Project description of Reflections of HAL and Samantha online. Link:https://tillmannohm.com/reflections-of-hal-and-samantha/

[9] | Zarkadakēs (2016)

belief systems everything contains a 'soul' or 'spirit': animals, plants, rocks, rivers, the weather and maybe even words. George Zarkadakis argues that symbolic language predates the modern mind, and is not only used to communicate, but ultimately dictates how the world is represented in our own consciousness. Only through language we place 'souls' and 'spirits' into everything around us and art objects fulfill the function to mirror our own minds physically and make connections with the world around us. It is no surprise then that the original artist of the lion-man created a hybrid of a human and a mountain lion, probably imagining themselves into the figure of the predator and using the object as a symbol to communicate ideas.

In the intro of "In Our Own Image" George Zarkadakis tells the story of how the fictional character Robby the Robot from the movie *Forbidden Planet* impressed him as child so much, that he would imagine the robot as a playmate that could walk, talk and obey orders. When he later decided to study engineering and exchanged thoughts on artificial intelligence with other researchers, they all seemed to relate to the same sci-fi movies, books and stories. The hint here is that these fictional stories influence what is getting built and how we think about the future.

Inspired by this revelation I asked participants at the *Silicon Friend Camp* about their first contacts with a robot. My own memories brought me to a picture of myself, proudly holding a plastic toy robot in my hand. The doll—or 'action figure'—had a battery compartment, which was the power source for tiny wheels underneath it's feet and light bulbs that it had for eyes. It could not talk, but I remember that it made noises that resemble guns or a laser sounds. In the 80s and 90s the bleak vision of boxy humanoid robots carrying guns was very prominent. Movies like *Terminator* and *RoboCop* fueled my young male fantasy for power and destruction.

The other first encounters were less militaristic. One artist remembered a robot dog that he got as a present. In the early 2000s digital pets with integrated sensors became cheap enough for consumers. *Tekno the Robotic Puppy*[10] by Manley Toy Quest was one of the more affordable robodogs in the market and with light and proximity sensors, buttons and even a microphone it was well equipped to learn some 'tricks' by triggering the right combination of sensors. Another artist was sharing an image of the *Tamagotchi* and while it was not a robot, the LCD screen in an egg-like plastic shell displayed a simulation of a pet. With 3 buttons children could interact with the virtual pet and fulfill its needs for food, affection and sleep. The anthropologist Pat Shipman proposes that animal connection can be considered a human trait that has uniquely evolved in us, next to tool making, symbolic language and the domestication of animals and plants.[11] Simulating animal connection by creating simple programs on microchips was then the next logical step in our drive to create symbols all the way down and fulfill the need of children to take care of others.

[10]The robot puppy was featured on the TIME magazine cover with the title "Tech comes to Toyland." Link: http://content.time.com/time/covers/europe/0,16641,20001211,00.html

[11]| Shipman (2010)

Yet another popular toy that was shared by an artist in our group and filled the same niche of robotic pets was *Furby*. This furry ambiguous creature with big eyes was made to simulate language acquisition and brings us back to the topic of how dolls learned to speak. In the case of *Furby* the designers decided to create a new vocal language called 'furbish.' When the device was first started it only repeated gibberish sentences and slowly over time would exchange these words with the local language equivalents as it ages. The original source code for one of the microchips was released in 2018[12] and even though I can not read assembly code, it is thoroughly documented through comments. The game logic of the furby.asm file is fairly simple, jumping between subroutines depending on the state of the many sensors, triggers and timers and occasionally using a pseudorandom number generator to mix things up. The actual speech sounds were accessed from a memory unit and send to a cheap Texas Instruments speech synthesizer chip. These chips have been in use for some decades already, most prominently in another toy called *Speak and Spell*, but also in watches, clocks and translators. The synthesis uses a prerecorded audio signal that is then compressed using linear predictive coding (LPC). This type of encoding allowed the voice to be later changed in frequency, pitch and loudness. But just like the *Vocoder*, it only digitized a voice and was not able to construct new words or sentences. So the original *Furby* from 1998 could only express a couple hundred words that were recorded by a voice actor, which is again only a digital equivalent to Chatty Cathy.

More powerful computers were already able to generate any kind of text into somewhat intelligible speech. The *Software Automatic Mouth (S.A.M.)* from 1982 was one of the first commercial products that used rule based formant synthesis. Instead of using whole prerecorded words, it uses an array of computer generated phonemes and places them together to make up sentences. While this was often understandable, it was not perfect. It sounds unnatural and robotic and was reliant on converting text to speech in a probabilistic manner, often mispronouncing words.[13] *S.A.M.* did not need any special hardware and could run on Apple, Atari and Commodore computers at the time, even though that would often mean using every CPU cycle and holding other programs. The software had its biggest moment at the launch event of the Macintosh computer in 1984, greeting the audience with a live demo in first person.[14] Sam and it's multiple software variations became the universal voice of the computer. The net.art pioneer Alexei Shulgin took the speech synthesis capabilities similar to *S.A.M.* to the extreme with his cyberpunk rockband *386 DX*.[15] The band is named after the Intel chip inside of

[12] Sean Riddle requested the Furby source code from the US patent office and uploaded it on his webpage. Link: http://www.seanriddle.com/furbysource.pdf

[13] | Lukose and Upadhya (2017)

[14] To make the demo work, engineers had to use a prototype Mac that was more powerful than the retail version. It said: "Hello, I am Macintosh. It sure is great to get out of that bag! Unaccustomed as I am to public speaking, I'd like to share with you a maxim I thought of the first time I met an IBM mainframe: Never trust a computer that you can't lift! Obviously, I can talk, but right now I'd like to sit back and listen. So it is with considerable pride that I introduce a man who has been like a father to me… Steve Jobs!" Link: https://www.folklore.org/StoryView.py?project=Macintosh&story=Intro_Demo.txt

[15] The songs of 386 DX are available on Alexei Shulgin's website. Link: http://www.easylife.org/386dx/

the old office computer, which is also the front singer and instrumentalist. It uses MIDI and voice synthesis to cover popular songs by The Doors, Nirvana or the Sex Pistols, traveling to festivals around the world and sometimes performing on the streets. The artist Alexei Shulgin takes the role of the operator, only pressing the play button and creating the visuals on the screen during the live performances. In my opinion this ironic take of making the microchip into it's own one-computer-band was a demonstration of how terrible and reductive rock music becomes, when it was compressed into bits and bytes. At the same time the novelty and humor of this process, generated enough interest for *386 DX* to produce 2 CDs and distribute them over music labels.

The strangely robotic voices that got associated with computers improved over time and when Apple's digital assistant Siri came out in 2011 they used a different process of concatenating audio samples together. Before Apple even started working on their voice assistant, a company called Scansoft auditioned hundreds of voice-over artists in 2005 and made them speak a multitude of random sentences. The idea of how to use the random utterances for new outputs was quite simple: cut the voice into smaller pieces of syllables, demisyllables, phonemes, diaphones or triphones and put them back together for any word you need. Doing this manually was—of course—very time consuming, so that programs were developed to analyze and categorize the speech dataset based on their acoustic properties. Then, at runtime an algorithm selects the best sequence to create the desired phrase. The results in the early 2000s still contained a lot of error and needed a substantial amount of tweaking, but the outcome was perceived as much more natural(REF). In 2006 Scansoft merged with Nuance, another company working on enterprise speech solutions, who presumably used the database to create the voice for Siri and licensed it to Apple. The original voice-actor in the US was Susan Bennet[16], who was oblivious of the fact that suddenly, 6 years after recording, her voice appeared on iPhones all over the country.

The technology to generate new sentences from previously recorded voices drastically improved, when deep weighted networks started to be used to analyze and select audio samples. One major achievement in generating natural voices came in 2016 in the form of a model named *WaveNet*[17] from a team of researchers at Google's DeepMind. They used convolutional node networks (CNN) to perform statistical analysis on raw audio files with 16000 samples per second. Using convolutional operations on temporal audio data meant that the researcher did not need to convert it first into a compressed format using a vocoder and the results were much more natural sounding than any previous text-to-speech algorithm. Using a raw audio dataset also means that the system can find patterns in any audio signal, including music. For demonstration the researchers generated new samples from a classical piano dataset, which sounds fascinating, but has more resemblance to free jazz than a classical piano score. This also happens to speech, when the network is conditioned without any text sequences it generates impressive

[16]Interview by the Guardian with multiple voice actors for Siri. Link: https://www.theguardian.com/technology/2015/aug/12/siri-real-voices-apple-ios-assistant-jon-briggs-susan-bennett-karen-jacobsen

[17]Oord et al. (2016)

human babbling where real words are interspersed with made-up word-like sounds and breathing noises. When the network is trained on audio and the transcription together the system does not only generate a consistent voice replica of a single human, but can also be used for voice recognition tasks transforming utterances back to text.

Other companies experimented with weighted networks to better create concatenative voice models. At the same year in 2016 Adobe showcased a hypothetical feature for their audio software, where users can create voice clones with a small 20-minute dataset and then use text to edit spoken words in a sentence. In the presentation project *VoCo* at the Adobe MAX conference, the developer Zeyu Jin used an example sentence and exchanged the words in such a way that instead of "I kissed my dogs and my wife," the audio was manipulated to say "I kissed Jordan three times" – referring Jordan Peele, the host of the conference. This strange demonstration, that was supposed to be funny, got a lot of attention online and created a debate around the ethical implications of manipulated audio and voice cloning. Adobe never released *VoCo* as a product, maybe because of the media backlash and fear from the public, but other companies like *Lyrebird*[18] filled the gap, trying to market their product for the use of editing podcasts and voice-overs. However, they started their media campaign by cloning the voices of Donald Trump and Barack Obama to "discuss" about the startups product and praise it for their democratizing effort of making it available to anyone.[19] Together with synthetic video, so called "Deepfakes" have earned a bad reputation for being used in fraud[20] and pornography[21]. At the same time artists are using the technique to create hypothetical stories to mock politicians and other famous figures.[22] Zeyu Jin, the creator of *VoCo*, already proposed watermarking synthetic voices and use software to detect if a voice is fraudulent.

We can see that with the release of *WaveNet* and *VoCo* in 2016 a new field in voice synthesis opened up. Hybrid systems were created that used both concatenative and generative techniques to categorize and interpolate between tiny audio samples of human speech.[23] The first release of *WaveNet* was computationally expensive, but already one year later Google implemented a new version that works in real-time for their voice assistant and has shown that a slim hybrid model can be used on low-powered devices, such as smartphones[24] and digital assistants today are moving towards the use of synthetic voices generated with deep weighted networks.[25]

[18] Lyrebird changed their name to Descript. Link: https://www.descript.com/

[19] | *Politicians discussing about Lyrebird* (n.d.)

[20] Apparently fraudsters were able to use a synthetic voice to instruct a bank transfer of 220.000€. (Stupp 2019)

[21] In 2019 an app with the name DeepNude got public attention due to a Vice article and was taken down by the author shortly after. Cole (2019)

[22] Examples of deepfake political comedy can be found on the YouTube Channel of Speaking of AI. Link: https://www.youtube.com/channel/UCID5qusrF32kSj-oSGq3rJg

[23] | Pierce (n.d.)

[24] | "WaveNet launches in the Google Assistant" (n.d.)

[25] | Y. Wang et al. (2017)

Conversational Agents

For the Amazon founder Jeff Bezos, the board computer in Star Trek was the inspiration for investing in the cloud-based voice software Alexa.[26] The product was initially marketed as a networked speaker. The software is now expanding to other items, including watches, smartphones, jewelry, light bulbs and doorbells. The aggressive price war with Amazon-connected products is partly to capture the connected home market, but also to collect as much natural voice data as possible. Over the past 10 years, deep weighted networks[27] have become popular for classification and pattern recognition tasks and have especially made an impact in automated speech recognition. Before weighted networks were used, the error rate of automated speech recognition (ASR) went from 40% in the 1990s down to 15% in the 2000s, which was still very high, but became practical to use. Back then the user would have to feed the system with voice samples, so that it could reliably turn utterances into text, but even then they would need to speak like a robot to be understood. With a landmark paper[28] in 2012 by research groups at the University of Toronto, Microsoft, Google and IBM it became that the application of deep weighted networks outperforms any of the previous techniques. In the same year Android released it's new cloud based speech recognition service[29] and since then users can speak into their phone naturally and be transcribed quite accurately. For a long time though, the models needed gigabytes of data in computer memory, to create accurate results, so that users need to send their voice samples to large server farms for transcription. Only in the past 2-3 years offline recognition became feasible through specialized hardware and model optimization, trading performance for robustness. Google's claim that new speech recognition systems have an error rate of less than 5% is only true for a specific—mostly white and male—audience speaking American English and drops abruptly with a more diverse distribution of dialects, age and gender.[30]

In order for deep ASR systems to perform well they need a massive amount of input data. A paper from the Chinese tech giant Baidu proposes 12000 hours of speech to create a robust model.[31]bA global internal database of speech recordings enables the company to improve its speech recognition. The intrusion into privacy is immense and has already been used by U.S. law enforcement agencies as evidence in a court case.[32] But there is

[26]Hydra was developed by Olivia Jack. Link: https://hydra.ojack.xyz

[27]As described in the introduction I use "weighted" instead of "neural."

[28]| Hinton et al. (2012)

[29]Google AI Blog article about "Speech Recognition and Deep Learning." Link: https://ai.googleblog.com/2012/08/speech-recognition-and-deep-learning.html

[30]Tatman 2017, did a study on Gender and Dialect Bias in YouTube's Automatic Captions that shows a higher error rate for women and non-American speakers. Facebook's AI research department created a dataset of people with different skin tones, gender and age to evaluate speech recognition models for biases and found significant differences between lighter and darker skin tones (Liu et al. 2021).

[31]Ref: Amodei et al. (2015)

[32]The first Alexa recording in a court case was handed over after the defendant agreed handing over his data. The Independent, Amazon Echo could become key witness in murder investigation after data turned over to police, 2017. Link:https://www.independent.co.uk/news/world/americas/amazon-echo-murder-investigation-data-police-a7621261.html

another, more ethical, way than stealing people's voices: The Mozilla Foundation's *Commonvoice* project relies on people voluntarily recording their voices for computer models, and the resulting speech recognition and synthesis software can be offered with an open source license.[33] Common speech and text datasets already have been publicly available from projects like VoxForge that uses the same crowdsourced techniques as Mozilla or LibriSpeech, which essentially scraped and aligned public domain audiobooks.[34] Other common non-free resources come from the Linguistic Data Consortium, who, for example, created *Switchboard*, a corpus of conversations between two US citizens randomly connected and recorded over the telephone line. Listening to the audio sample provided on the website creates oddly voyeuristic feelings. We can hear a mundain conversation of a man talking to a woman about his interest for gardening.[35]

Artist Lauren Lee McCarthy plays with these tensions between intimacy and privacy, convenience and agency. In her projects *LAUREN* and *SOMEONE* she installs connected devices into volunteers homes and either acts as a control system herself or lets others remotely monitor the volunteers and control the devices in their homes. This creates an interesting tension, when the person knows that there is an actual human listening and watching from afar. At the same time the artist and performers find themselves in a position of a helpful voyeur.

However, the role of human labor behind voice assistants is not just about executing and understanding commands. People tend to interpret voices and categorize them according to age, gender and social status. Companies take advantage of this and design their voice software according to certain identity schemes, which are provided with a history, hobbies and preferences. It is precisely this illusion that excites users and makes the product interesting. In a UNESCO think piece titled *I'd blush if I could*[36], they explore harmful gender biases associated with digital assistants. The voice assistants of major tech companies are scripted by default as female personas with smart, humble, and sometimes funny personalities. The teams working on voice assistants try to avoid this aspect, because Apple, Microsoft and Google ask their employees to refer to their headless voices as "it" and when the persona get's asked the question directly winds out with a joke. Only Alexa answers what is obvious in the default design of all of them with "I'm female in character." By being submissive, they thus support a patriarchal image of female stereotypes that we already know, in a historical context, from human computers and other secretarial roles. The paper calls for women to be more empowered and involved in IT. It calls for AI software to avoid gender attributes whenever possible, and for AI assistants to take a clear stand against sexist behaviors.

Researcher and artist Nadine Lessio creates useless voice assistants to critique the current corporate agenda of productivity, efficiency, and consumption. She does this by using the programming interfaces provided to make apps for corporate voice assistants.

[33] Link:https://commonvoice.mozilla.org/de
[34] Link to LibriVox: https://librivox.org/ and Panayotov et al. (2015)
[35] Link to the LDC Switchboard-1 Release 2: https://catalog.ldc.upenn.edu/LDC97S62
[36] The title "I'd blush if I could" is also the response Siri gives to the insult "You're a bitch"

For example, she explores the concept of a depressed home assistant with *SAD Home (Depressed Alexa 1.0)*[37], an Alexa hack that grants users their wishes depending on the weather and other mood factors, sometimes it simply turns itself off.

This scripted denial of a capitalist logic ironically uses the same technique as the company behind it. Voice assistants are carefully crafted by a team of creative professionals working in the field of "conversation design." In the book *Talk to me*[38] the Author James Vlahos describes that many people in the field are far away from computer science and more commonly had careers in the liberal arts. The teams are made up of authors, playwrights, comedians, actors as well as anthropologists, psychologists, and philosophers who imagine the personality of the AI persona that should represent the brand. To create the character, they have to come up with all possible questions and create various answers for each of them. Vahos recalls asking Microsoft's Cortana "Where do you come from?" and the female voice replies "I was made by minds across the planet."[39] And even though the designers decided to use the first person "I," they really are talking of themselves carefully crafting the answers played back by the loudspeakers around the world.

The tedious process of mapping out all questions and creating answers for them is mostly done through creative writing, but also utilizes careful statistical analysis of the questions users send to the cloud.

It is interesting how the term AI is used in the context of voice computer interfaces, because there is nothing "smart" about it, just a winding flowchart of if-else conditions. What might be clever is the nefarious way of how companies trick people into the belief of computer personalities, encouraging people to interact with the device like children playing with dolls. And when a company like Microsoft has experimented with a more sophisticated chatbot, like Tay[40], it started to repeat the racist and misogynistic slurs of twitter users and consequently has to be heavily filtered.

Leewa

The artist duo *Ekheo* consisting of Aude Langlois and Belinda Sykora explore the crosslines of musical, visual and sonic art performances. They met at the Sound

[37] Nadine Alessio's project website. Link: http://nadinelessio.com/projects.html

[38] | Vlahos (2020)

[39] | Vlahos (2020) p.117

[40] Tay was the name of a chat bot Microsoft intended to have a teenage personality and could be interacted with over multiple channels. Twitter trolls co-opted the bot with sexist, racist and antisemitic questions that the bot replied to with generic answers. They also made use of a repeat-after-me phrase to make it look like the bot is spewing hateful comments itself. After only 16 hours Microsoft deleted all accounts and the PR disaster was immense, with headlines happily personifying the AI that has learned to be racist. But no continuous learning algorithm was involved. Link to a blog post by Russel Cameron Thomas explaining the technology stack behind Tay. ("Exploring Possibility Space: Microsoft #TAYFAIL Smoking Gun: ALICE Open Source AI Library and AIML" n.d.)

Studies program at the Universität der Künste in Berlin, where they created musical improvisations and theoretical frameworks on voice and music. They started their research on voice, gender and identity in relation to AI systems in 2019 and held talks at the MUK[41] and other institutions to share their insights. At the *Silicon Friend Camp* they wanted to extend their already existing improvisation and research practice by integrating a form of embodied AI. But the main focus was the voice itself: How is gender perceived through it? How do we navigate the tension between what is considered an artificial voice and an authentic voice? To tackle these issues during the camp they wanted to create and record soundscapes based on vocal improvisation and use AI systems in the form of synthetic voices, generative composition tools, robotic elements and so on. So we prepared the attic of the house with a small sound studio that they could work in and during the week they found inspiration in using a multitude of AI systems and human collaborators that they used to create an improvisation and finally a 6-minute video and audio composition. To come up with the identity of their AI partner, they used the *Oracle*[42] to ask questions of their name, gender, favorite instrument etc. which reads like this:

> You: how old is she?
> Oracle: Today she will be 11.
> You: what is her favourite musical instrument?
> Oracle: her greatest pleasure is to whisper to the sounds of trumpets and the beating of tympani
> You: is she happy?
> Oracle: It is a mystery and a great sadness.
> You: What is Leewa's surname?
> Oracle: It was suggested that her human's name was /u/Seattle_Celtics.
> You: does she like to sing?
> Oracle: I have no voice
> You: No I mean Leewa
> Oracle: Query : Have you ever cheated?
> You: Does Leewa have a voice?
> Oracle: Yoow!
> You: what is Leewa's favourite poem?
> Oracle: Leaves of Grass
> You: What is Leewa's favourite book?
> Oracle: Leewa does joke around and loves to give sarcastic responses. She also loves to read books, especially horror novels.
> You: What gender does Leewa have?
> Oracle: She is female.
> You: which music does Leewa like?
> Oracle: Leewa likes to sing and most of all Leewa likes to dance.

[41] | "GENDER & DIVERSITY: Gastvortrag von Ekheo - Musik und Kunst Privatuniversität der Stadt Wien" (n.d.)
[42] The Oracle was a terminal computer connected to GPT-3 with a generic Q&A prompt.

> You: What is art for her?
> Oracle: art is the process of iterating through consciousness generation.

This served as a base for the identity of *Leewa* and inspired further experiments using a whispering voice and 3D printing a female doll bust. To recreate the voice I searched for tools to do voice synthesis and found an open source toolbox from Corentin Jemine that claimed to do Real-Time Voice Cloning from a few seconds of audio.[43] The method is essentially based on a previous paper from Google engineers: *Transfer Learning from Speaker Verification to Multispeaker Text-To-Speech Synthesis*[44] (SV2TTS), which itself uses a pipeline of a speaker-encoder network (GE2E) to represent different voices in different positions of the latent space, a text to waveform synthesizer (Tacotron) and a vocoder to generate sound from these waveforms (WaveNet).[45] As described previously combining different encoding and generator systems into a larger framework is commonly used and is commercially used in Google's Assistant and their cloud services for other companies.

Thanks to the efforts of Corentin Jemine and other contributors, who built a user interface to explore these tools, we were able to input our own voices and generate new samples with pre-conditioned voice models. Unfortunately though, the results were not as expected, by inputting only a small number of samples the output was not picking up on our voice characteristics and the results were too noisy and of low-quality. Instead we tried out the cloud service from resemble AI[46] that is mentioned in the code repository and where some of the maintainers work. Instead of the intended use of cloning one individual voice, we decided to create a collective voice of the camp participants. Every one of us recorded 10 phrases that were used as an input to the model.

After I pressed "train" on the web interface and waited for some time for the process to finish, I got an e-mail telling me that the "Resemble score" of our voice was 53%. So the transfer learning process was not quite successful, but the resulting voice was very surprising. I expected that the algorithm would try to average our voices into something that fits in between all data points. Instead it created clones of a voice for some phonemes, but not for others, so that a sentence randomly interpolates between different speakers. It also seems that, at least in this run, the characteristics of male voices are more prominent than female voices, but in general it has a strange "artificial" nature to it, where the pitch and speed of the voice is constantly changing.

Because we used a commercial product, we did not have access to the local model and could only interface with it through their web interface and APIs, which made using the voice for real-time performances impossible. Instead *Ekheo* created some voice samples

[43] From the github repository CorentinJ/Real-Time-Voice-Cloning. Link: (Corentin Jemine [2019] 2022). and his master thesis: (C. Jemine, n.d.)
[44] | Jia et al. (2019)
[45] The SV2TTS pipeline is based on these previous papers: Wan et al. (2020) → Y. Wang et al. (2017) → Oord et al. (2016)
[46] | "Clone Synthetic AI Voices with Neural Text to Speech" (n.d.)

that they played live during the final performance. For this, they also recorded real voices from participants at the camp and intertwined it with synthetic drones and other instruments. The visuals were created from a live video feed of the doll, that was modified by Naoto Hieda using the live coding shader language hydra[47] The puppet itself could move it's head, which was remotely controlled by So Kanno via a servomotor.

In my opinion the example of *Leewa* illustrates the collective efforts that go into building "an AI" quite nicely: it shows the process of using open source and proprietary code developed by people reliant on previous research, the creation of collective datasets through the recording of our voices and finally the human performers hiding in the background and remotely manipulating symbols for the audience. At the same time *Leewa* is a self-referential project that makes the collective experience throughout the camp audible, by using our synthetic and real voices, the interaction with the *Oracle* and snippets of anecdotes from the camp.

---

[47]Hydra was developed by Olivia Jack. Link: https://hydra.ojack.xyz

Piles of Data

Self-Learning Networks

The idea behind Artificial Neural Networks has a long standing history. Using our understanding of the brain as a blueprint for mathematical operations dates back to the 1950s when the psychologist Frank Rosenblatt developed the *perceptron*.[1] Inspired by nerve cells and their connections (synapses), the perceptron takes multiple input values, sums them up and outputs a 0 or 1 depending if a predefined threshold is reached. This system can be 'trained' by using positive and negative reinforcement to define the weights for each connection. With an apparatus, the Mark-I Perceptron,[2] that uses photoreceptors to turn light into 'bits' (today we would say pixels) the perceptron could 'sense' shapes in the form a binary matrix and distinguish between circles, squares and triangles. He proposed that a network of perceptrons could possibly even recognize faces and objects. Even though he developed the perceptron in 1964, Frank Rosenblatt never got to see his invention take off.[3] Another engineer, Kunihiko Fukushima, kept refining his methods in the 70s by adding multiple layers, effectively creating the first 'deep neural network' where deep just means the depth of 'hidden' or inbetween layers connecting the input signal to the output classifier.[4] He called this self-organizing system Cognitron[5] which was successful at accurately recognizing numbers and letters on a 12x12 grid. It's successor the Neocognitron[6] took further inspiration from the visual cortex and a discoveries by Hubel & Wiesel made in the 1950s that some biological neurons selectively respond to local features like lines, edges or color and others to figures like circles, squares or even human faces. This is also the core idea behind convolutional neural networks (known as ConvNet or CNN) which separate an image into a smaller grid and apply a certain filter to them, e.g. checking for edges. The french computer scientist Yann LeCun came up with ConvNets in the 1980s which are *the* driving force for AI systems today. Additionally Geoff Hinton, a cognitive psychologist and computer scientist, popularized

[1] (Rosenblatt 1958)

[2] The Mark I was a electromechanical machine that used motor driven potentiometers to adjust the variable weights.

[3] Frank Rosenblatt died in a boating accident in 1971. A couple years prior Marvin Minsky heavily criticized the mathematics behind perceptrons and advocated for a symbolic approach. These turn of events might have lead to a lack of funding in the 'connectionist' AI research field and ultimately lead to a general disinterest when the symbolic approach could not keep their exaggerated promises.

[4] (Mitchell 2019), p. 114

[5] (Fukushima 1975)

[6] (Fukushima 1988)

the backpropagation algorithm in 1986 which finally made it possible for filters to tune themselves instead of relying on predefined rules.

Most conceptual ideas behind current deep weighted networks were already present in Frank Rosenblatt papers,[7] but weighted networks were often outperformed by rule-based systems. So what changed when Alex Krizhevsky, a student of Hinton, made a phenomenal leap in 2012 on the ImageNet classification competition? The main innovation is outlined by Krizhevsky in the paper itself: "To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation."[8] Until AlexNet was released it was incredibly time consuming to run the conditioning process on the CPU which can only do one operation of matrix multiplication at the time. The GPU as the name Graphics Processing Unit suggests was originally designed to calculate 3D scenes and render them on a display. This involves a lot of matrix and vector operations and to accelerate them GPUs are capable at calculating large blocks of data in parallel. This means that conditioning AlexNet on 1.2 million pictures took only 6-9 days on two consumer graphics cards compared to probably weeks or months without them. However it was not the first system that utilized the GPU, it is similar to a CNN by Dan C. Cire an et al. released a year prior which has a reported speedup of 60 times for large weighted networks.[9] The other roadblock for deep weighted networks is 'overfitting.' In the case of the ImageNet competition that would mean that the model adapts to the image so closely that it would simply reproduce the categories of the image without being able to identify new pictures which were not inside the training set. The most common way to reduce overfitting is to have a sufficiently large dataset with a high amount of variance. For AlexNet the 1000 classes and 1.2 million images were still not enough and they used data augmentation which transforms, flips or changes the color of an image to increase the training set by a factor of 2048. This means that—in theory—a larger dataset increases the robustness of the weighted network.

In conclusion, the conceptual framework of how weighted networks (or artificial neural nets, or perceptrons, or cognitrons) work was mostly established in the past century. Their performance today comes down to increased computing power and larger training sets. In fact datasets have become bigger and bigger and we haven't seen the limits of what weighted networks are capable of, they seem to be mainly limited by data and computation. In the following chapter I will have a closer look at how large training sets are constructed in an academic and artistic context—both my own and a few other examples—and the ethical issues and responsibilities regarding privacy, consent, representation and ownership.

[7] (Tappert 2019)
[8] (Krizhevsky, Sutskever, and Hinton 2017)
[9] "One epoch takes 35 GPU minutes but more than 35 CPU hours." (Ciresan et al., n.d.)

Making Data, 11076 hands

My first encounter with a large dataset in the field of computer vision was in late 2017 when I found *11k Hands*.[10] As the name suggests it is a dataset of 11076 human hands compiled by the researcher Mahmoud Afifi for gender recognition and biometric identification. The images show the hands of 190 individuals, in both their palm and dorsal orientation, placed on a white background with uniform lighting. Each image is accompanied by the following metadata: age, binary gender, skin color on a scale of "very fair," "fair," "medium" and "dark," left or right hand, palm or dorsal, if it has nail polish and if there are "irregularities." A statistical analysis of the dataset shows that it contains more female than male hands, mainly people in there 20s and a majority of "medium" and "fair" skin tones. The gender bias is addressed in the paper and mitigated by filtering the training set to have an equal amount of males and females. They report that the CNN conditioned on this dataset had on average a 87% accuracy recognizing the correct gender on an image of the palm and a 90% accuracy for the dorsal side. But it is not the 'state-of-the-art' results that drew me to the paper, it was the gender and skin-tone classification itself that appalled me. It reminds me of phrenology in the 19th and 20th century, a popular pseudoscience that claimed that persons character and mental abilities could be determined by the shape of their skull. Phrenologists like Franz Joseph Gall went to a great length measuring and categorizing human skulls and associating certain regions to human traits. While it didn't take scientists to debunk the idea that bumps on a head could indicate the characteristics of a person, it had a comeback in the early 20th century when it was applied to racist and sexist stereotypes and used to justify Nazi eugenics. The underlying assumption that the shape of the head has anything to do with the mental state of the person was simply wrong.[11] Considering that gender is a social construct and not necessarily a binary choice trying to use a computer to identify if a person is male or female by analyzing their hand seems arbitrary at best and reinforcing gender stereotypes at least.[12] Another aspect of the dataset stood out to me when I started to look through the images. These were not pictures in the traditional sense, their aesthetical value did not matter as they are simply a tool to to accomplish a task. They are not made 'to look at' instead these images of hands are produced for a computer to analyze. Harun Faroki called these types of images 'operative' in his three-part series *Machine/Eye* where he examined how military technologies like guided weapons produce images that serve only the utility of the machine. As Aud Sissel Hoel puts it: "operative images are devoid of social intent, that they are not meant for edification, and nor for contemplation."[13] But what happens when you contemplate on the hand dataset was remarkable to me. When I looked through the images quick

[10](Afifi 2018). Data and source code available here: (Afifi n.d.)

[11]In 2018 Parker et al jokingly tested the Galls theory using 21st century scientific methods and MRI data. (Dempsey-Jones 2018)

[12]I didn't do any testing of the system as I don't know how to run MatLab code, but I can imagine that the slightly better results on the dorsal hand are the result of nail polish only applied on female hands.

[13](Hoel 2018)

enough I could not only see the motion of the test subject as the images were selected frames from a video, but I also started to imagine the person behind the hand and apply my own stereotypes and biases that far expanded the labels of the dataset. I could easily imagine the scene around the camera with a couple researchers who assembled a makeshift photo studio in the office of their lab. They greet the subject with a handshake, which is probably a fellow student, and explain the procedure quickly to get them to sign the paper that their anonymized personal data will be published for scientific research. Then the person puts both their hands under the camera spread their fingers and leave. This dichotomy between the label and my own narrative inspired me to create a video piece featuring the unaltered dataset. I used the default mac computer speech synthesis to read out the labels corresponding to each hand, and sped it up to fit into a 26-minute-long video. As viewers witness the participants holding their hands into the camera and spreading their fingers, they are met with a monotone, beat-like soundtrack of repeating words like "fair" and "medium" and occasionally "nail polish."[14] Example of a frame from 11k hands

At the 36th Chaos Communication Congress I organized a workshop called "Palm Reading AI" where I invited visitors to read the hands of people from the 11k hands dataset. At first I only introduced them to palmistry using wikihow as a reference.[15] Then I handed out some prints where a random hand from the dataset was depicted and participants had to fill in a couple of questions. Some questions were short guesses like age, gender, country of origin, for some other they had to come up with fictional stories with only the hand lines as a reference: what is the persons future? How was their childhood? How is their love life? After they filled in the form some people shared their stories and then I revealed where these hands came from and how computer scientists are using them to create models that try to predict their gender. Afterwards we had a discussion about the practice of creating large datasets and their ethical considerations. I had a longer talk with one participant that did not want to guess the age or gender of that person and I had told them that this was exactly the point of the workshop: to reflect on our own biases and stereotypes and how they translate into science.

After long contemplation on 11k hands and finding datasets that are much more problematic than this one, I don't think the type of work from Afifi et al. is 'unethical' or needs to be redacted. They got consent from their subjects and share the dataset to the scientific community for "reasonable academic fair use." The work on biometric identification and comparing CNNs to previous methods is interesting and novel, however as I stated above I think the premise behind the gender recognition task is flawed. Unfortunately this is very common in the computer vision field where people are (mis-)labeled that reflect and amplify societal stereotypes.[16] My research on the hands dataset in

[14] See: (Matthias Schäfer 2018)

[15] ("How to Read Palms: 9 Steps (with Pictures)" n.d.)

[16] One particular famous example of this is the work by Michael Kosinski and Yiluna Wang. Their flawed study tried to predict if a person is gay by scraping dating sites and training a classifier on these images. See:(Murphy 2017)

conjunction with esoteric practices and fortune-telling informed a later work of mine "The Chiromancer" that I built together with Giacomo Piazzi.

In the wild

The process of collecting and creating data has drastically changed since the wide adoption of the internet. Where the 11k hands dataset has invited participants to their instute to take a picture specifically for the dataset, other researcher started to search and download huge collections from the internet without any consent. Take for example the ImageNet dataset, initiated by Stanford University's AI professor Fei Fei Li, which was created to tackle object recognition tasks and eventually consisted of 14 million images.[17] The team queried multiple search engines with common nouns and multiple translations to get around 500-1000 images per category. The categories are derived from an older project called WordNet, created in 1985 at Princeton University, that tried to achieve a hierarchical ontology of words. For example the noun *human* is synonym to *homo, man and human being* which are subcategories of the *hominid* class, which are *primates*, which are *mammals*, which are *animals*, which is an *organism* and finally part of the *entity* class. After downloading millions of images in thousands of categories they were first automatically deduplicated but the task of checking whether the image actually depicts the search term could not be done by an algorithm. Originally Fei Fei Li wanted to hire graduate students to sort the data which was too costly and time consuming.[18] One of her undergraduates suggested to use a new service called Amazon Mechanical Turk where people around the world complete small task for little compensation. Even with the help of Mechanical Turk it took 2,5 years to sort and validate the first dataset with 3.2 million images in over 5000 categories. At some point ImageNet was the largest academic client for Amazon's Turk service and after the popularization of weighted networks they became a corner stone of data annotation. Similar platforms were established and Mechanical Turk grew to an estimate of half a million workers, which were rigorously exploited and alienated as a so called click workers. One paper estimates their median hourly wage to 2$ and only the top 4% earning more than 7.25$ per hour.[19] If the task is rejected by the requester the worker does not get paid at all and has no way of appealing that decision. To advocate for better working conditions the designer and researcher Caroline Sinders built an open source tool for data annotation and a wage calculator to better estimate the cost of labeling.[20]
With so many different people working on labeling images, societal biases are eventually reflected in the dataset. In 2019 the artist Trevor Paglen and researcher Kate Crawford collaborated on an exhibition titled *Training Humans*, dedicated to human image

[17] ImageNet started with 3,2 million images and had the goal to collect 50 million by the end of 2011. (Deng et al. 2009)
[18] (Gershgorn 2017)
[19] (Hara et al. 2017)
[20] See: ("Technically Responsible Knowledge" n.d.)

datasets. One of the main exhibits was a vast collection of human images from ImageNet.[21]
As you go further down the category tree you might find images of people that fall into categories such as "Bad Person, Call Girl, Drug Addict, Closet Queen, Convict" and so on.[22] The artists used these absurd, racist, and misogynistic labels to train *ImageNet Roulette*,[23] a recognition algorithm that was accessible online and in an interactive installation. People online quickly picked up on the tool and shared images of themselves with sometimes amusing and often deeply problematic image captions.[24] While some of the people in the research community defended ImageNet that these offensive labels are not part of competition and make up only a small fraction of the total dataset, as a result of the media attention that followed more than 600,000 images were removed. It is now only accessible after proving to be part of a scientific institute.

Another pair of artists and researchers, Adam Harvey and Jules LaPlace, have been exposing image sets which often get revoked and removed after publishing their investigation.[25] Harvey and LaPlace focus on datasets with faces that are captured "in the wild." One particular example was the *UnConstrained College Students (UCCS)* dataset captured at the University of Colorado Colorado Springs. According to the authors of the associated papers to the dataset they identified that current image sets created for face recognition research did not address the presence of unknown subjects. The authors of the papers associated with this dataset wanted to create a more realistic benchmark for face recognition research by introducing unknown subjects over time. To achieve this, they captured students on campus during breaks using a 800mm telephoto lense from a distance of 100-150 meters through an office window. The authors frame it as a benefit that the students are unaware of their capturing for the dataset as they casually focus on random activities and their faces are sometimes blurred and occluded. To make things worse this research was mainly funded by US defense and intelligence agencies. The public backlash was immense and as a result the dataset is no longer publicly available. The researchers in question did not apologize and argued that their subjects are anonymous as their names or other identifiers are not published.[26] However when researchers need unconstrained and non-consensual data, they do not often capture them directly using creepy surveillance tactics. Starting with "Labeled Faces in the wild" in 2007 the practice of collecting and labeling image data from internet sources has become normalized and is still largely unregulated. Often these image sets "in the wild" operate in a gray zone where they either depict public figures, arguing that privacy regulations do not apply to them, or that they simply link to the file and only store them temporarily for analysis. A third option uses media licensed under Creative Commons (CC) which is mostly considered free and legal data with no restrictions in the AI research

[21] See: ("KATE CRAWFORD | TREVOR PAGLEN: TRAINING HUMANS – Fondazione Prada" n.d.)
[22] ("Excavating AI" n.d.)
[23] See: ("ImageNet Roulette – Trevor Paglen" n.d.)
[24] (Rea 2019)
[25] See: (Harvey n.d.a)
[26] ("2nd Unconstrained Face Detection and Open Set Recognition Challenge, ECCV-2018" n.d.)

community.[27] In a detailed report Adam Harvey lists many datasets that exploit and often misuse CC licenses to create face and object recognition datasets.[28] He identifies Flickr as a major source for collecting images, a popular image sharing website where users can choose from various Creative Commons, copyright and public domain licenses. Flickr actively promoted the use of CC licenses and offered unlimited free hosting if a CC license is used. Their strategy worked and by 2022 they amassed 467 million CC licensed images. In 2014 a joint research group including Yahoo Labs, the company that bought Flickr, shared one of the largest public media datasets[29] with the name *Yahoo! Flickr Creative Commons 100 Million (YFCC100M)*. A multitude of specified datasets were created from this corpus. One of them is the *MegaFace* face recognition dataset with 4,7 million faces from 672,057 identities. While all of the images fall under a CC license most of them prohibit their commercial use and require appropriate attribution, which was not given in the dataset. As Harvey and LaPlace verified, the *MegaFace* dataset was used by globally by commercial, military and academic organizations.[30] As an investigative article from the New York Times explains, most people are unaware that their images are powering face recognition research around the world, including companies with ties to Chinas surveillance on the Uyghur population.[31] To conclude the three common issues with CC licensed media is that they are not or wrongly attributed, the use of non-consensual biometric data is prohibited in some places (e.g. Illinois) and the use in commercial applications is often prohibited. Many of thes issues identified by Harvey and LaPlace remain until today and operate in a legal grey zone. While the notion behind uploading images under an open license for others to freely share, distribute and remix media is a noble goal, the CC license is legally weak and practically useless against opting out of statistical analysis in AI research. But we might be slowly moving into a direction where law makers catch up and create precedences that disallow the use of biometric data in certain applications, like the upcoming AI Act.[32]

This Person Does Exist

A dataset using Creative Commons images that I examined more closely was *Flickr Faces HQ (FFHQ)*. In 2018 the research lab of Nvidia, one of the leading companies for visual computing, published a paper introducing a machine learning architecture called StyleGAN.[33] They improved on generative adversarial networks (GAN) in such a way that it was possible to create controllable synthetic high-resolution images. In simple

[27] See a longer analysis on the exploitation of CC media by A. Harvey: ("Creative Commons Biometrics" n.d.)
[28] ("Creative Commons Biometrics" n.d.)
[29] YFCC100M only contains links and metadata to images and videos under CC license onf Flickr
[30] (Harvey n.d.b)
[31] (Hill and Krolik 2019)
[32] The Open Future Foundation is a think tank that actively tries to influence european digital policy debates. See ("Open Future – Open Future Foundation" n.d.)
[33] (Karras, Laine, and Aila 2019)

terms the system is able to abstract large amounts of images with a model that in turn outputs similar-looking pictures. In this case the model is able to generate realistic-looking photographs of human faces. In comparison to other datasets *FFHQ* is fairly small with only 70,000 images. As existing datasets were too low in resolution a new corpus was created by scraping Creative Commons, Public Domain or U.S. Government Works licensed images through Flickr's API. The dataset itself is published under a CC-BY-NC-SA license and the instructions for use and download are very clear, making it manageable for me in terms of size and effort to discover the underlying characteristics. The dataset consists only of photographic images as it was checked by Amazon Mechanical Turk workers to make sure that statues, paintings, or photos of photos were removed. Using the open-source library dlib, the raw images were automatically aligned and cropped around the face to create a uniform square ratio of 1024px. This library also identifies 68 points outlining the chin, eyebrows, eyes, nose and mouth which are included in the metadata. So the final dataset used for training the GAN consists of a multitude of faces whe the eye and mouth positions are always in the same spot.

Screenshot of This Person Does Exist

One year after Nvidia released its StyleGAN paper the software developer Phillip Wang published the website *thispersondoesnotexist.com* which shows the capabilities of the generative model to create realistic looking photographs of people.[34] The site quickly took off and alarmed people about potential impact of AI systems in generating cheap synthetic media.
As a counter-narrative to the AI image creator, I wanted to showcase the people who were used to train this system. In 2020 I moved the cropped and aligned face images to my server and built my own website with the name *this-person-does-exist.com* which displays the faces from the training set alongside their metadata.[35] Looking at the individuals faces facilitates an interpersonal connection with the unknown person and evokes a feeling for the images that were used to train the generative model. At the same time it shows the creepy and strange practice of AI researchers using personal images as raw data. As Flickr is used mostly by hobbyists and professional photographers, one can find portraits of children and families, speakers at conferences or people on holiday. We can assume that, like other scraped datasets, the photographers of these images are unaware that their images are used for AI analysis. In contrast to other scraped datasets, the Nvidia researchers provide a tool to see if an image is part of the collection and allow the removal of the photograph.[36] According to Adam Harvey the company does not disclose if any images were requested for removal and has not updated or removed any photos since its release in 2016.[37]

70k faces from FFHQ compiled into one image The Flickr Face - averaged FFHQ

[34] At the time of this writing the url redirects to stability.ai, an archived version can be found. See ("This Person Does Not Exist" 2019)
[35] See (Schäfer 2021)
[36] ("FFHQ Dataset Search Form" n.d.)
[37] See 'FFHQ Dataset' at ("Creative Commons Biometrics" n.d.)

dataset

To get a sense of scale of the dataset, I compiled all face images into a grid, reducing the size of each image to 16 by 16px. This simple montage makes it possible to get a feeling for the vast amount of normalized image data. The authors claim that FFHQ includes more variety than other face image sets in terms "of age, ethnicity and image background, and also has much better coverage of accessories such as eyeglasses, sun-glasses, hats, etc."[38] They admit to biases inherited from the Flickr platform but fail to mention them. Looking through the data it becomes clear that a majority of the images are taken with high resolution digital camera systems under good lighting conditions. Many of the pictures show separate the subject with a blurry background, so called 'bokeh,' and conform to photographic aesthetic norms of the early 2000s until today. Indeed, the dataset contains a variety of skin tones, age groups and background but they are not equally distributed. To visually find a bias in image sets I averaged the pixel values of the images. This suppresses outliers, but it allows us to see an overall trend of the dataset. The resulting composite image The Flickr Face, I belief, reveals a trend towards smiling and light-skinned people in the data set. All these are highly subjective impressions, but I would not presume any ethnicity or categorize people by skin tone. I understand that this project is doing the same as the research labs in question, but I hope by making people uncomfortable and showing the human behind AI systems, we can get a better feeling for how creepy this harvesting of faces is.

As a part of a growing group of artists exploring and exposing research datasets, I coined a term for this genre: Dataset Art. My paper on this subject was published in *Temes de Disseny* in 2021 and includes a couple more examples. Through their works, these artists are able to make large datasets understandable and captivating. Their art has sometimes even created enough attention to lead some institutions to remove questionable parts or entire collections. Whether through galleries or online, Dataset Art is providing an exciting new way to peek into the workings of AI systems.[39]

Scrapism

Web scraping is the technique of using computer programs to automatically visit links and aggregate data from the internet. It is the backbone for many of the current machine learning applications. The artist Sam Lavigne uses the practice of web scraping with goal of creating art with an emotional or critical message; a practice he calls 'Scrapism.'[40] Perhaps an early work using the internet as a data source to convey emotion is *Listening Post* by Mark Hansen and Paul Ruben. The work, that got awarded the Golden Nica in 2004, uses snippets from internet chat rooms and displays them on over 200 LED signs while a crude computer voice reads them out loud. The artists are transforming live

[38] (Karras, Laine, and Aila 2019)
[39] (Schäfer 2021)
[40] (Lavigne n.d.)

data into a sensual experience where the viewer gets to enter a dark room and listen in to the worlds chatter happening simultaneously all over the planet.[41] Instead of using and exposing datasets made for scientific research, Lavigne creates his own datasets by downloading and analyzing materials on the internet to revert common power structures. For example the online art work *New York Apartment* he produced together with Tega Brain for the Artport Collection of the Whitney Museum.[42] In this work they collected all for-sale listings on multiple real estate websites for New York City apartments and created a website that compiles all of them into one giant apartment listing. The value for this fictional apartment is over 43 billion USD and boasts 65,674 bedrooms and 55,588 bathrooms on around 3.4 million square meters. The website consists of multiple columns describing the listing with all of the clichéd language and staged photographs common in the real estate market. They extruded the floor plans into 3D models and placed them together next to each other, in a tower or pyramid formation so that you can explore this maze of apartments. The videos are cut up into thematic categories like "Welcome," "Bedroom," "Master" or "Pre-war" creating strange super cuts of panning and zooming shots of slick interiors. Although looking through this compilation is funny and entertaining, it reveals the absurd language of luxury commodities and reveals the inequality of who can afford to own housing in a city like New York. Another experiment by Sam Lavigne uses an open source hair detector to create a compilation of Mark Zuckerberg hair styles, in reaction to multiple people mocking the billionaire for his 'terrible' hair style.[43] This project in particular surved me as an inspiration as it shows that using web scraping and machine learning technologies can sometimes just be for silly projects on the expense of billionaires funding this tech for surveillance and personal profits.

Sam Lavigne - Zuckerhair

hidden.pictures

When downloading things from the web, we often assume that everything is stored for ertenity. "The web never forgets" is a common phrase that is used to warn people before uploading sensitive and personal content online. This can be true for content that is widely shared or automatically scraped and stored on sites like archive.org, but for many files and links the web is brittle. Domain names expire and cut the link to the requested page. But even if the hypertext can be accessed they often contain links to files that do not exist on the server anymore, in the case of images the browser then decides to show a broken image icon and an empty rectangle. For the online artwork *hidden.pictures* (previously called *empty.photos*) I created a web crawler that visits random URLs and whenever it hits an image that did not load correctly it gets stored into a database. I

[41] See ("X-TRA → Mark Hansen and Ben Rubin: Listening Post" n.d.) and ("Ars Electronica Archiv" n.d.)
[42] (Lavigne and Brain n.d.)
[43] (Sam Lavigne 2020)

collected thousands of broken image links together with their metadata that sometimes describes the image.[44] I present the viewer with a collage of randomly placed image rectangles on a blank white website. A visitor can pan around in all directions finding the default rendering of their web browers for broken images, often showcasing a ripped paper icon on the top left corner. When hovering over one of the boxes the original url is shown on the bottom left corner of the page and whenever it existed the alt-tag pops up next to the cursor. These images from thousands of blogs, shops and forums show the forgotten and neglected part of the internet. It invites for the imagination about the internet that is not getting maintained and reveals an even bigger source of data that does not exist anymore. When looking at this through the lense of weighted networks and training data, we have to ask the question what these models are actually learning when so many things get missing on the net every day. On the other hand the models themselves become a blurry snapshot of the things they were able to gather, but might not exist in the future.

empty.photos exhibited at Best Off 2019

doggg.art

In 2020 I created doggg.art, an exercise in *scrapism* where I downloaded and transformed content from the social media giant Instagram. Instagram has become the biggest tool for artists to find an audience and a community. Every possible niche can be found through the use hashtags like #dogart that collects drawings, photographs of personal and commercial dog related imagery. With around a million posts it is only one example for the immense creative output on the image platform. Facebook, the mother company of Instagram, owns an abundant amount of image data which they analyze and use to optimize computational models. By posting on the platform a user "grant[s] a non-exclusive, royalty-free, transferable, sub-licensable, worldwide license to host, use, distribute, modify, run, copy, publicly perform or display, translate, and create derivative works of [their] content."[45] I would describe doggg.art as a generative big dada collage consisting of over 30k images from Instagram tagged with #dogart. The images were processed using a machine learning algorithm called U^2-Net, which removed the background and other elements from the pictures. A website then randomly places the cutouts on a beige background slowly fading them in, creating an ever changing digital dog meadow. I see it as a collaborative work that combines pieces by 38326 artists together credited on a seperate page with all their unique usernames.[46] The artwork was intended as a screensaver that reflects on on the aesthetics of the platform Instagram and how our relationship with pets extends into this online network.

doggg.art exhibited at Die Digitale Düsseldorf exhibition "Digital Jokes"

[44] See (Schäfer n.d.)
[45] See ("Nutzungsbedingungen | Instagram-Hilfebereich" n.d.)
[46] Complete list of artists can be found here https://doggg.art/artistlist.html

A visitor looking through the names of dog artists

A study on the Characteristics of Douyin / Meanwhile in China

Working with and about social networks has been a big part in my artistic work. In 2019 I was invited by UNESCO to participate in a residency in the city of Changsha, the capital of the Hunan region and "City of Media Arts" in central China. While I was there I tried to make sense of the information landscape around me and got hooked to the chinese clone of TikTok. Douyin became one of the most successful apps worldwide as the leading platform for creating and sharing short videos. Created by the Beijing based company Bytedance, it is one of the few Apps that got successful outside of the great firewall. To comply with chinese law Douyin is a completely separate App from TikTok. Even though the interface and logo looks the same, the content is completely different, not accessible from the international version. I asked S()fia Braga to join me as a collaborator, inspired by the work of *I stalk myself more than I should* where she captures Instagram stories, a feature designed to expose a video for a limited duration of 24 hours, and reappropriates them in a video installation. The aim of our work was to explore and analyze the vast digital ecosystem of Douyin from different perspectives, using their recommendation algorithm to lead us to different aspects of chinese social media. Using screen recordings to capture hours of video footage of us scrolling through our feeds. These found images are then decontextualized without alteration to give visitors the space to reflect upon them, gain insight into a walled off platform and into the algorithms designed for user engagement.
We made two video installation, running an 8 hours of captured material, showing people dancing next to chinese police forces schowcasing their equipment. Sometimes we would show the blank interface showing blank search results. Search terms like Donald Trump result in videos that do not show the american president them self, indicating facial recognition being used for censorship. Additionally to the unaltered screen captured we created non-nonsensical graphs that invoke the feeling that the content is used for statistical analysis. The title *A study on the characteristics of Douyin* was taken from a sociological paper and gives the veneer of scientific legitimacy to the non-consensual stalking of capturing of our practice. In another exhibition at the Ars Electronica Festival in Linz the titled was changed to *Meanwhile in China* and the wall behind the video screen was covered with a collage of graphs and datapoints found online showcasing the exploding growth and user distribution on the platform. However the graphs themselves are stripped of any labeling, making them unreadable and useless as visualization.

A Study on the Characteristics of Douyin at Xie Zielong Photography Museum in Changsha

Meanwhile In China at Ars Electronica 2019

Recommended Hashtags & More of the same

I have had an account on Instagram since 2012 shortly after it was released on android devices. The social network was based around square images, often highly edited with filters that are inspired by old analog photos. When I joined Instagram I already had a critical view of large social networks like facebook where I had started to create online and offline performances. On Instagram I would often share references to technologies in the social sphere and where they break, like screenshots of loading spinners, weird advertisement or comments by bots. To get the attention outside of my friend network and reach even more bots I often used generic hashtags. The hashtag originated from the IRC protocol where # was used before the name of a chat room (channel) and became a popular way on twitter for tagging events, movements or topics from 2007 onwards.[47] Today it is a ubiquitous way of organizing and tagging content into categories, but it is mostly used to get attention inside of the network. For this, the game on social media is to add as many relevant and trending topics into the description as possible and many companies have sprung up to help users 'optimize' their hashtags. From 2015 I started to use hashtags for non-sensical and humorous image describtion and later on decided to use hashtag generators to create lists of irrelevant tags for my posts on Instagram as social commentary. In 2020 just after the first summer of the new pandemic reality S()fia Braga organized the *Internet Yami Ichi*, a black market for internet inspired products invented by the japanese art collective IDPW.[48] For this flea market I decided to print and sell the 5 years of my Instagram performance with a book titled *Recommended Hashtags*. Each of the 216 pages containing the image, hashtags and number of likes. It also had a dedication to all my followers at the end page. After downloading and printing my Instagram data, I deleted all my images on the platform and added the label 'Post-Instagram Artist' to my account. As users of Facebook's platforms, we are aware that our content is being analyzed for profit. With my online performance, I aimed to subvert and add noise to the system, albeit not enough to make a real difference. Selling my book was a way to compensate for the 'work' I have been doing for Facebook. The book was sold in a staggering model where each edition doubled in cost from the previous one, making my old data more valuable over time, while it was completely free at the time of creation. In the following year 2021 I started a new performance titled *more of the same* on my account using the dataset I downloaded as source material for a generative adversarial network. I trained a StyleGAN model, which was originally invented to create synthetic human faces, on the images I posted from 2015 to 2020. The model however created abstract shapes, colors and textures instead of replicating my images. As ~200 pictures are a very small dataset and the images are not very uniform ranging from screenshots to photographs depicting objects or people the model could not converge on any meaningful representation. Many of the generated images are nearly symmetrical, reminiscent of Rorschach inkblots, which is the result of augmenting my image dataset by flipping the images. They also show circular blobs which is the result of the model

[47] See (Chakelian 2021)
[48] See ("The Internet Yami-Ichi" n.d.)

architecture and is seen as an error of the weighted network.[49] In the end I wrote a small program that uploaded and posted a new generated image with the caption "more of the same" for 100 days.

Screenshot of more of the same | 200

Book cover of Recommended Hashtags | 200 First page of Recommended Hashtags | 200

[49] See (Karras et al. 2020)

Stochastic Meaning

```
What I had not realized is that extremely short exposures
to a relatively simple computer program could induce powerful de-
lusional thinking in quite normal people.[^1]
```

Symbolic Language

Understanding language is an essential aspect for computer assistants to be truly useful. However, language is a challenging area since words can have various meanings depending on the context. Furthermore, there are multiple languages and different ways of converting spoken language into written symbols.

The use of computers for natural language processing (NLP) has been a subject of research since the invention of computers. One of the earliest examples of symbolic NLP is ELIZA, a computer program developed by Joseph Weizenbaum in the 1960s that simulated a conversation by rephrasing the input to form questions. ELIZA used pattern matching and substitution to generate responses based on the user's input, giving the impression of understanding what was being said. For example, if the user said "I feel sad," ELIZA might respond with "Why do you feel sad?" or "Tell me more about your feelings."

While ELIZA was fairly limited in its capabilities, it sparked the interest and imagination of the general public which saw the toy program as a breakthrough in artificial intelligence. Many people were fascinated by the idea that a machine could understand and respond to human language, even if the responses were simple and scripted. Joseph Weizenbaum himself was shocked about the willingness of people to attribute consciousness to even the most simple programs, like the 200 lines of code he wrote for the DOCTOR program that simulated a Regorian psychotherapist. In a famous anecdote Weizenbaum explains that his office assistant, who was quite familiar with the program, demanded for him to leave the office because she was sharing intimate thoughts with the computer. Some psychologists even considered using such a computer program to be able to treat more people in a shorter time, which for Weizenbaum was a horrendous thought and finally made him doubt the future of humanity and computers. In his 1976 book "Computer Power and Human Reason," Weizenbaum argues that the increasing reliance on computers to solve problems and make decisions is a dangerous trend that could lead to dehumanization and loss of autonomy of human beings. Conflating the

computer with the human mind ultimately reduces the human to a computational being. He warns that we should not confuse the power of machines with human intelligence, creativity, and empathy, which are essential for ethical and meaningful interactions.[1]

ELIZA paved the way for further development in natural language interaction. In the 1970s some programmers created text-based games like the game *Colossal Cave Adventure* (1976) which allowed users to interact with the game world using natural language commands. At the same time the use of conceptual ontologies, which are structured representations of knowledge in a particular domain became popular. These ontologies can be used to help computers understand natural language by mapping words and phrases to their corresponding concepts in the ontology. For example, if someone says "I want to buy a red car," an ontology could map "red" and "car" to their corresponding concepts and infer that the person is interested in purchasing a vehicle with a specific color. One of the biggest endeavors in mapping words to concepts was the creation of WordNet in the 1980s, which is a lexical database of English words and their semantic relationships.

Even until the early 2000s programmers came up with larger and more complex rule-based chatbots. Creating systems like A.L.I.C.E. (1995) which used complex set of rules to simulate conversation with users. However, these systems were limited by their reliance on pre-programmed responses and lacked the ability to learn and adapt to new situations. The web application Cleverbot (2008) and it's predecessor jabberwacky (started in 1988, online since 1997) shifted static databases to one that is constantly growing.[2] By capturing every conversation and creating links of question-answer pairs, the system could use a fuzzy search algorithm to present the user with a likely response that was typed by another human previously.

While symbolic NLP methods have been successful in some applications (such as chatbots and question-answering systems), they have limitations when it comes to handling ambiguity and understanding human emotions or nuances in language. As a result, more recent approaches to NLP have focused on machine learning techniques that allow computers to learn from large amounts of data without relying on pre-defined rules or ontologies.

Attention is all you need

- Prayers to GPT ## Computing Reality
- The Chiromancer

[1] Ibid.

[2] See ("Jabberwacky - About Thoughts - An Artificial Intelligence AI Chatbot, Chatterbot or Chatterbox, Learning AI, Database, Dynamic - Models Way Humans Learn - Simulate Natural Human Chat - Interesting, Humorous, Entertaining" 2006) & ("Cleverbot" n.d.)

Thank you to my supervisors *Manuela Naveau*, *Christa Sommerer* and *Laurent Mignoneau*. A special thanks to *Davide Bevilacqua* and the *Servus.at* association for helping me realize the Silicon Friend Camp, exhibition and symposium.

Thank you to *Bani Budassin* who invited me to write for Temes des Disseny which formed my thoughts around Dataset Art.

And of course to all my new and old friends, my fellow contemporaries. I am because you are.

(Infrastructure???)

- Infields / Cloud
- Cybersquat

"2nd Unconstrained Face Detection and Open Set Recognition Challenge, ECCV-2018." n.d. Accessed March 11, 2023. https://vast.uccs.edu/Opensetface/.

Afifi, Mahmoud. 2018. "11k Hands: Gender Recognition and Biometric Identification Using a Large Dataset of Hand Images." September 16, 2018. https://doi.org/10.48550/arXiv.1711.04322.

———. n.d. "11k Hands." Accessed March 7, 2023. https://sites.google.com/view/11khands.

Amodei, Dario, Rishita Anubhai, Eric Battenberg, Carl Case, Jared Casper, Bryan Catanzaro, Jingdong Chen, et al. 2015. "Deep Speech 2: End-to-End Speech Recognition in English and Mandarin." http://arxiv.org/abs/1512.02595.

"Anatomy of an AI System." 2018. Anatomy of an AI System. 2018. http://www.anatomyof.ai.

"Ars Electronica Archiv." n.d. Accessed March 20, 2023. https://archive.aec.at/prix/showmode/88/.

"Avatars." n.d. So Kanno. Accessed April 28, 2022. https://www.kanno.so/project/avatars.

"Babycastles." n.d. Babycastles. Accessed January 30, 2022. https://www.babycastles.com.

Bailey, Jason. n.d. "DeepDream Creator Unveils Very First Images After Three Years — Artnome." Accessed March 4, 2023. https://www.artnome.com/news/2018/12/30/deepdream-creator-unveils-very-first-images-after-three-years.

"Better Images of AI." n.d. Accessed March 4, 2023. https://betterimagesofai.org.

Bostrom, Nick. 2001. "Are You Living in a Computer Simulation?" 14. https://www.simulation-argument.com/simulation.pdf.

Chakelian, Anoosh. 2021. "The Journey and Many Faces of the Hash Symbol." New Statesman. June 9, 2021. https://www.newstatesman.com/science-tech/2014/06/history-journey-and-many-faces-hash-symbol.

Ciresan, Dan C, Ueli Meier, Jonathan Masci, Luca M Gambardella, and Jurgen Schmidhuber. n.d. "Flexible, High Performance Convolutional Neural Networks for Image Classification."

"Cleverbot." n.d. Cleverbot. Accessed March 21, 2023. http://www.cleverbot.com/.

"Clone Synthetic AI Voices with Neural Text to Speech." n.d. Resemble AI. Accessed January 26, 2022. https://www.resemble.ai/.

Cole, Samantha. 2019. "This Horrifying App Undresses a Photo of Any Woman With a Single Click." Vice. June 26, 2019. https://www.vice.com/en/article/kzm59x/deepnude-app-creates-fake-nudes-of-any-woman.

"Creative Commons Biometrics." n.d. Accessed March 11, 2023. https://adam.harvey.studio/creative-commons/.

Dempsey-Jones, Harriet. 2018. "Neuroscientists Put the Dubious Theory of 'Phrenology' Through Rigorous Testing for the First Time." The Conversation. January 22, 2018. http://theconversation.com/neuroscientists-put-the-dubious-theory-of-phrenology-through-rigorous-testing-for-the-first-time-88291.

Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. 2009. "ImageNet: A Large-Scale Hierarchical Image Database." In *Cvpr09*.

"Descartes Was Wrong: 'A Person Is a Person Through Other Persons' | Aeon Ideas." 2017. Aeon. April 7, 2017. https://aeon.co/ideas/descartes-was-wrong-a-person-is-a-person-through-other-persons.

"Everything — DAVID OREILLY • Computer Art & Research." n.d. Accessed April 28, 2022. https://www.davidoreilly.com/everything/.

"Excavating AI." n.d. -. Accessed March 10, 2023. https://excavating.ai.

"Exploring Possibility Space: Microsoft #TAYFAIL Smoking Gun: ALICE Open Source AI Library and AIML." n.d. Accessed March 21, 2023. https://exploringpossibilityspace.blogspot.com/2016/03/microsoft-tayfail-smoking-gun-alice.html.

"FFHQ Dataset Search Form." n.d. Accessed March 15, 2023. https://nvlabs.github.io/ffhq-dataset/search/.

Francis Hunger. 2021. "1. 'Artificial Intelligence' => ' Automated Pattern Recognition' 2. 'Machine Learning' => 'Machine Conditioning' OR 'Automated Classification' 3. 'Neural Network' => 'Weighted Network' 4. 'Deep Learning' => 'Deep Conditioning' 5. 'Neuron' => 'Weight' or 'Node'." Tweet. Twitter. July 9, 2021. https://twitter.com/databaseculture/status/1413462059291975680.

"FREEPORT 1: ANATOMIES OF A BLACK BOX." n.d. Matadero Madrid. Accessed April 23, 2022. https://www.mataderomadrid.org/en/calls/freeport-1-anatomies-black-box.

Fukushima, Kunihiko. 1975. "Cognitron: A Self-Organizing Multilayered Neural Network." *Biological Cybernetics* 20 (3-4): 121–36. https://doi.org/10.1007/BF00342633.

———. 1988. "Neocognitron: A Hierarchical Neural Network Capable of Visual Pattern Recognition." *Neural Networks* 1 (2): 119–30. https://doi.org/10.1016/0893-6080(88)90014-7.

Gade, Christian B. N. 2012. "What Is *Ubuntu* ,? Different Interpretations Among South Africans of African Descent." *South African Journal of Philosophy* 31 (3): 484–503. https://doi.org/10.1080/02580136.2012.10751789.

"GENDER & DIVERSITY: Gastvortrag von Ekheo - Musik und Kunst Privatuniversität der Stadt Wien." n.d. Accessed January 24, 2022. https://muk.ac.at/veranstaltung/

gender-diversity-gastvortrag-von-ekheo.html.

Gershgorn, Dave. 2017. "The Data That Transformed AI Research—and Possibly the World." Quartz. July 26, 2017. https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/.

Hara, Kotaro, Abi Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch, and Jeffrey Bigham. 2017. "A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk." December 28, 2017. https://doi.org/10.48550/arXiv.1712.05796.

Harvey, Adam. n.d.a. "Exposing.ai." Exposing.ai. Accessed March 11, 2023. https://exposing.ai/.

———. n.d.b. "Exposing.ai: MegaFace." Exposing.ai. Accessed March 13, 2023. https://exposing.ai/datasets/megaface/.

Hill, Kashmir, and Aaron Krolik. 2019. "How Photos of Your Kids Are Powering Surveillance Technology." *The New York Times: Technology*, October 11, 2019. https://www.nytimes.com/interactive/2019/10/11/technology/flickr-facial-recognition.html.

Hinton, Geoffrey, Li Deng, Dong Yu, George Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, et al. 2012. "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups." *IEEE Signal Processing Magazine* 29 (6): 82–97. https://doi.org/10.1109/MSP.2012.2205597.

Hoel, Aud Sissel. 2018. "Operative Images. Inroads to a New Paradigm of Media Theory." In *Image – Action – Space*, 11–28. De Gruyter. https://doi.org/10.1515/9783110464979-002.

"How to Read Palms: 9 Steps (with Pictures)." n.d. wikiHow. Accessed March 8, 2023. https://www.wikihow.com/Read-Palms.

"Human-Guided Burrito Bots Raise Questions about the Future of Robo-Delivery." 2019. The Hustle. June 3, 2019. https://thehustle.co/kiwibots-autonomous-food-delivery/.

Hunger, Francis. 2021. "Talk: Unhype AI (2021)." April 28, 2021. https://www.irmielin.org/unhype-ai/.

"ImageNet Roulette – Trevor Paglen." n.d. Accessed March 11, 2023. https://paglen.studio/2020/04/29/imagenet-roulette/.

"ImageNet." n.d. Accessed March 4, 2023. https://image-net.org/index.php.

"Inceptionism: Going Deeper into Neural Networks." 2015. June 17, 2015. https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html.

"Jabberwacky - About Thoughts - An Artificial Intelligence AI Chatbot, Chatterbot or Chatterbox, Learning AI, Database, Dynamic - Models Way Humans Learn - Simulate Natural Human Chat - Interesting, Humorous, Entertaining." 2006. July 1, 2006. https://web.archive.org/web/20060701052309/http://www.jabberwacky.com/j2about.

Jemine, C. n.d. "Master Thesis : Automatic Multispeaker Voice Cloning," 38.

Jemine, Corentin. (2019) 2022. *Real-Time Voice Cloning*. https://github.com/CorentinJ/Real-Time-Voice-Cloning.

Jia, Ye, Yu Zhang, Ron J. Weiss, Quan Wang, Jonathan Shen, Fei Ren, Zhifeng Chen, et al. 2019. "Transfer Learning from Speaker Verification to Multispeaker Text-To-

Speech Synthesis." http://arxiv.org/abs/1806.04558.

Karras, Tero, Samuli Laine, and Timo Aila. 2019. "A Style-Based Generator Architecture for Generative Adversarial Networks." March 29, 2019. https://doi.org/10.48550/arXiv.1812.04948.

Karras, Tero, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. "Analyzing and Improving the Image Quality of StyleGAN." March 23, 2020. http://arxiv.org/abs/1912.04958.

"KATE CRAWFORD | TREVOR PAGLEN: TRAINING HUMANS – Fondazione Prada." n.d. Accessed March 10, 2023. https://www.fondazioneprada.org/project/training-humans/?lang=en.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2017. "ImageNet Classification with Deep Convolutional Neural Networks." *Communications of the ACM* 60 (6): 84–90. https://doi.org/10.1145/3065386.

Lavigne, Sam. n.d. "Scrapism." Accessed March 15, 2023. https://scrapism.lav.io/.

Lavigne, Sam, and Tega Brain. n.d. "New York Apartment." Accessed March 17, 2023. https://artport.whitney.org/commissions/new-york-apartment/index.html.

Lukose, Sneha, and Savitha S Upadhya. 2017. "Text to Speech Synthesizer-Formant Synthesis," 4.

Matthias Schäfer, dir. 2018. *11k Hands*. https://www.youtube.com/watch?v=snJsKFxPlJ8.

Miller, Arthur I. 2019. *The Artist in the Machine: The World of AI Powered Creativity*. Cambridge, Massachusetts: The MIT Press.

Mitchell, Melanie. 2019. *Artificial Intelligence: A Guide for Thinking Humans*. New York: Farrar, Straus and Giroux.

Molnar, Christoph. n.d. *10.1 Learned Features | Interpretable Machine Learning*. Accessed March 4, 2023. https://christophm.github.io/interpretable-ml-book/cnn-features.html.

Murphy, Heather. 2017. "Why Stanford Researchers Tried to Create a 'Gaydar' Machine." *The New York Times: Science*, October 9, 2017. https://www.nytimes.com/2017/10/09/science/stanford-sexual-orientation-study.html.

"Nutzungsbedingungen | Instagram-Hilfebereich." n.d. Accessed March 17, 2023. https://help.instagram.com/581066165581870.

Oord, Aaron van den, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. 2016. "WaveNet: A Generative Model for Raw Audio." http://arxiv.org/abs/1609.03499.

"Open Future – Open Future Foundation." n.d. Open Future. Accessed March 13, 2023. https://openfuture.eu.

Panayotov, Vassil, Guoguo Chen, Daniel Povey, and Sanjeev Khudanpur. 2015. "Librispeech: An ASR Corpus Based on Public Domain Audio Books." In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5206–10. South Brisbane, Queensland, Australia: IEEE. https://doi.org/10.1109/ICASSP.2015.7178964.

Pierce, David. n.d. "How Apple Finally Made Siri Sound More Human." *Wired*. Accessed January 23, 2022. https://www.wired.com/story/how-apple-finally-made-

siri-sound-more-human/.

Politicians discussing about Lyrebird. n.d. Accessed January 24, 2022. https://soundcloud.com/user-535691776/dialog.

Rea, Naomi. 2019. "How ImageNet Roulette, a Viral Art Project That Exposed Facial Recognition's Biases, Is Changing Minds About AI." Artnet News. September 23, 2019. https://news.artnet.com/art-world/imagenet-roulette-trevor-paglen-kate-crawford-1658305.

Rosenblatt, F. 1958. "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain." *Psychological Review* 65 (6): 386–408. https://doi.org/10.1037/h0042519.

"ruDALL-E." n.d. ruDALL-E. Accessed April 28, 2022. http://rudalle.ru/en/.

Sam Lavigne. 2020. "Just Discovered an Open Source Hair Detector and Have Used It on Hundreds of Images of Mark Zuckerberg to Build What I Believe Is the Most Comprehensive Archive of Zuckerberg Haircuts in Existence. Thank You AI Researchers! https://t.co/QJqXLAbqyw." Tweet. Twitter. February 16, 2020. https://twitter.com/sam_lavigne/status/1229097648818446336.

Schäfer, Matthias. 2021. "This Person Does Exist." *Temes de Disseny*, no. 37 (July): 214–25. https://doi.org/10.46467/TdD37.2021.214-225.

———. n.d. "Missing.pictures." Accessed March 17, 2023. https://missing.pictures/.

Shipman, Pat. 2010. "The Animal Connection and Human Evolution." *Current Anthropology* 51 (4): 519–38. https://doi.org/10.1086/653816.

Stupp, Catherine. 2019. "Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case." *Wall Street Journal: WSJ Pro*, August 30, 2019. https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402.

Tappert, Charles C. 2019. "Who Is the Father of Deep Learning?" In *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*, 343–48. Las Vegas, NV, USA: IEEE. https://doi.org/10.1109/CSCI49370.2019.00067.

"Technically Responsible Knowledge." n.d. Accessed March 10, 2023. http://trk.network/essay.

"The Canon Cat The Mac's Ancestor." 2014. October 25, 2014. https://web.archive.org/web/20141025053109/http://www.jagshouse.com/swyft.html.

"The Internet Yami-Ichi." n.d. In Kepler's Gardens. Accessed March 20, 2023. https://ars.electronica.art/keplersgardens/de/the-internet-yami-ichi/.

"This Person Does Not Exist." 2019. May 31, 2019. https://web.archive.org/web/20190531222303/https://thispersondoesnotexist.com/.

Tyka, Mike. n.d. "Deepdream/Inceptionism - Recap." Accessed March 4, 2023. http://mtyka.github.io/code/2015/07/21/one-month-after-deepdream.html.

Vlahos, James. 2020. *Talk to Me: Amazon, Google, Apple and the Race for Voice-Controlled AI*.

Wan, Li, Quan Wang, Alan Papir, and Ignacio Lopez Moreno. 2020. "Generalized End-to-End Loss for Speaker Verification." http://arxiv.org/abs/1710.10467.

Wang, Pei. 2019. "On Defining Artificial Intelligence." *Journal of Artificial General Intelligence* 10 (January): 1–37. https://doi.org/10.2478/jagi-2019-0002.

Wang, Yuxuan, R. J. Skerry-Ryan, Daisy Stanton, Yonghui Wu, Ron J. Weiss, Navdeep Jaitly, Zongheng Yang, et al. 2017. "Tacotron: Towards End-to-End Speech Synthesis." http://arxiv.org/abs/1703.10135.

"WaveNet launches in the Google Assistant." n.d. Deepmind. Accessed January 24, 2022. https://deepmind.com/blog/article/wavenet-launches-google-assistant.

Weibel, Peter. n.d. "AAA." www.kunstforum.de. Accessed January 16, 2022. https://www.kunstforum.de/artikel/aaa/.

"X-TRA → Mark Hansen and Ben Rubin: Listening Post." n.d. Accessed March 20, 2023. https://www.x-traonline.org/article/mark-hansen-and-ben-rubin-listening-post.

Zarkadakēs, Giōrgos. 2016. *In Our Own Image: Savior or Destroyer?: The History and Future of Artificial Intelligence*. First Pegasus books hardcover edition. New York, NY: Pegasus Books LLC.