

Künstliche Intelligenz als Wahrnehmungskrise: theoretische und ästhetische Herausforderungen

Rotary Club Erlangen, 7.3.2024



Von der „Good Old Fashioned AI“ zur selbstlernenden KI



1961

8

PROCEEDINGS OF THE IRE

January

Steps Toward Artificial Intelligence*

MARVIN MINSKY†, MEMBER, IRE

The work and applications of Marvin Minsky are appropriate to the-art. The library of processing both the g

Summary—The problems computers solve really differ in areas: Search, Pattern Induction.

A computer can do, in

A view of artificial intelligence

by FRED M. TONGE
University of California, Irvine

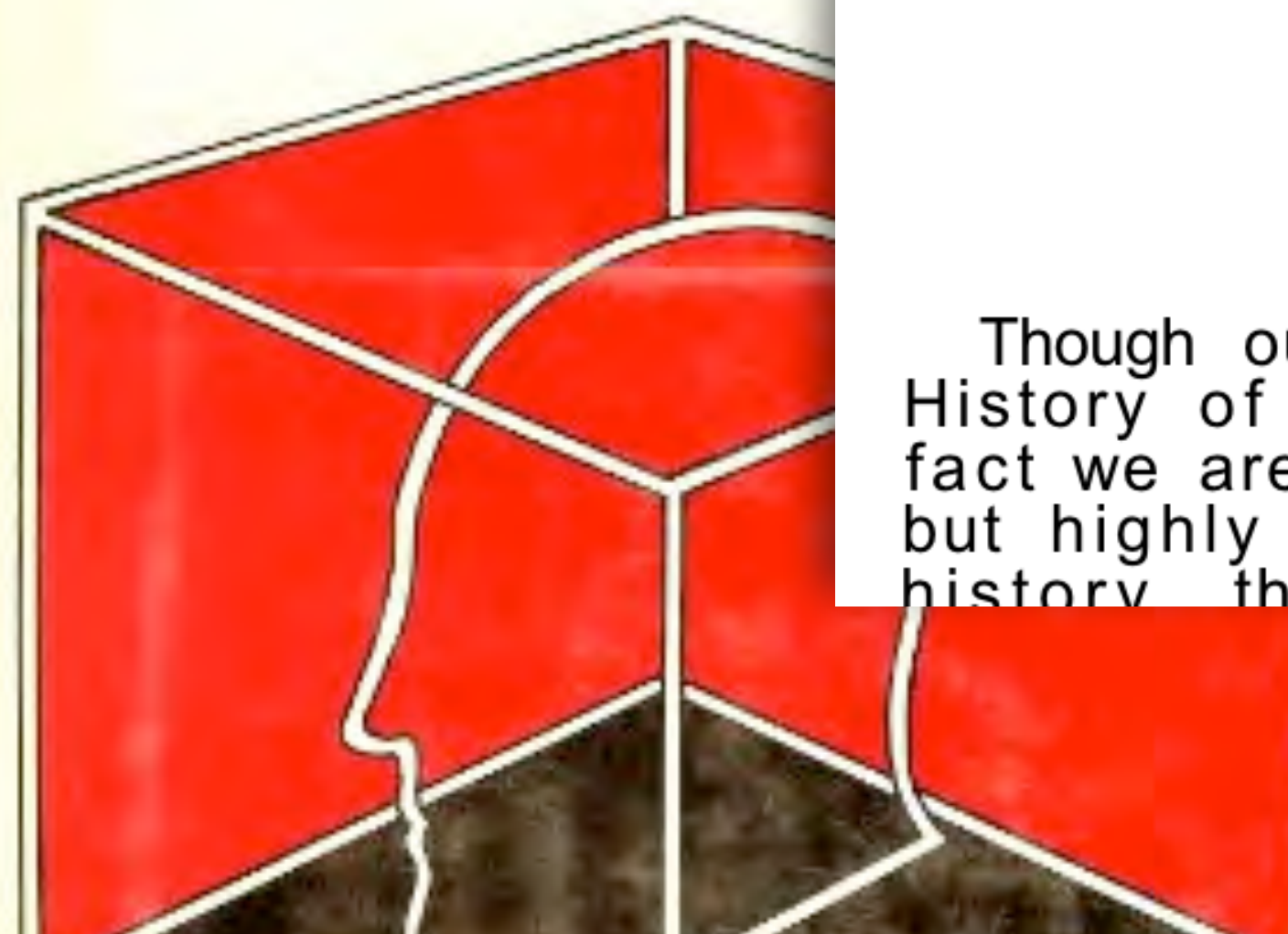
INTRODUCTION

By “intelligence” we mean a property of a system of behavior based on observation of the system’s behavior. It is agreed to by “most reasonable men” as intelligent. “Artificial intelligence” is then that property as it exists in non-living systems. Work directed toward producing such behavior is thereby work in artificial intelligence.

While the above is indeed a loose definition, it is useful in suggestiveness than in precision, it should serve for our purposes. It does contain at least one assumption — that, *a priori*, intelligence is not a property of “living” systems. And it does suggest that, if we are ever to know of whether artificial intelligence does or will ever exist is really worthy of further argument. If we can reach some agreement should be reached.

INTRODUCTION TO ARTIFICIAL INTELLIGENCE

PHILIP G. JACKSON



1966

1977

1974

HISTORY OF ARTIFICIAL INTELLIGENCE

P. McCorduck (Session Chairman), Univ. of Pittsburgh! M. Minsky, MIT: O. Selfridge, Bolt Beranek and Newman? H. A. Simon, Carnegie-Mellon University

The Early History

P. McCorduck

Though our discussion is entitled History of Artificial Intelligence, in fact we are focusing here on only a few but highly significant moments in the history — the moment when art

2004

PAMELA McCORDUCK
MACHINES WHO THINK



„Good Old Fashioned AI“ versus Deep Learning

- A. KRR: Wissenrepresentation und Schlussfolgern,
- B. PLAN: Planen
- C. MAS: Multi-Agent Systems
- D. RBT: Robotik
- E. PHIL: Philosophische Fragen
- F. NLP: Prozessieren natürlicher Sprachen
- G. CV: Computer vision
- H. ML: Machine Learning

12

1 From Logic to Cognitive Science

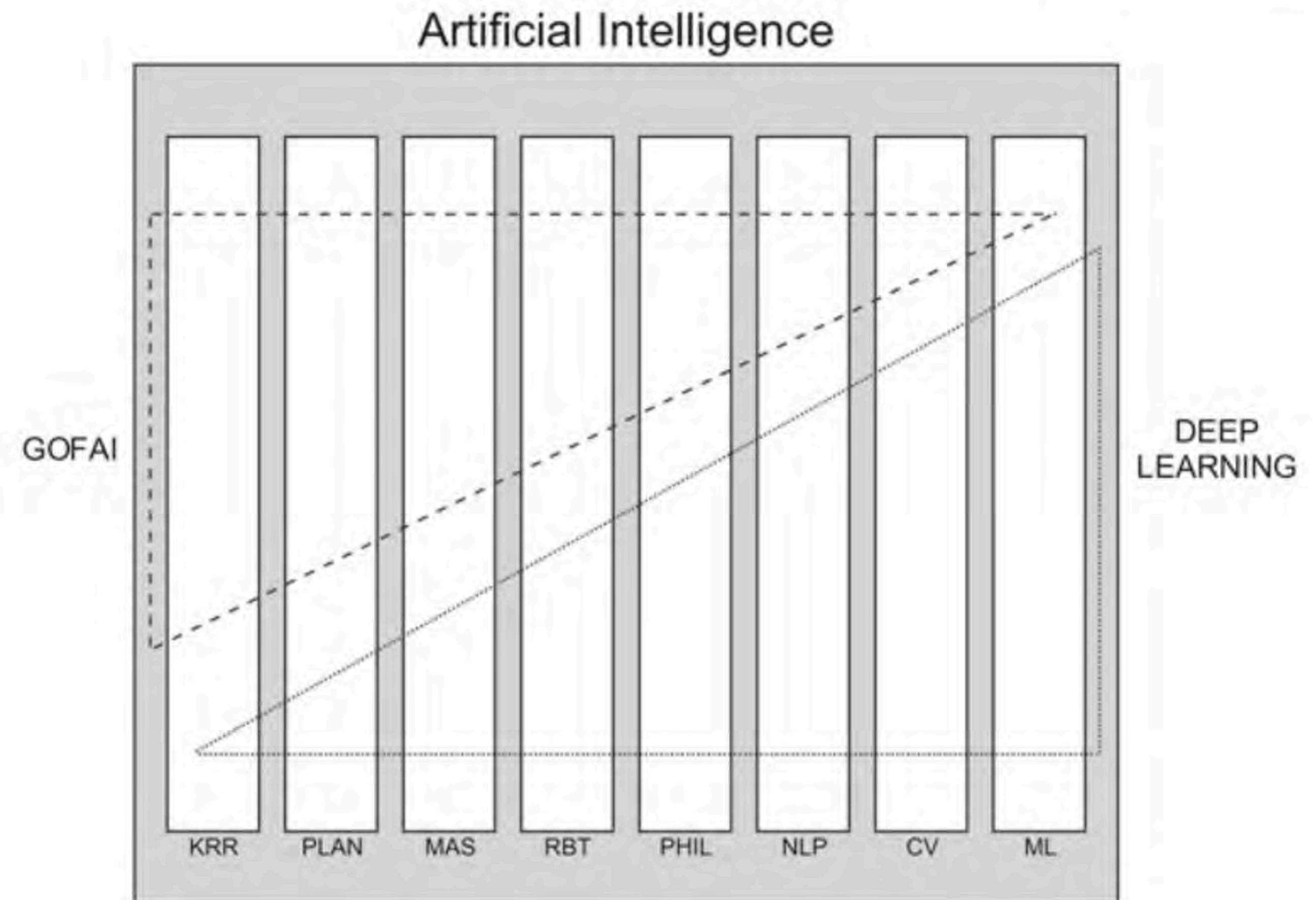


Fig. 1.1 Vertical and horizontal components of AI

GOF AI: „Logic-based Reasoning“

6.2 Reasoning with the Resolution Method

75

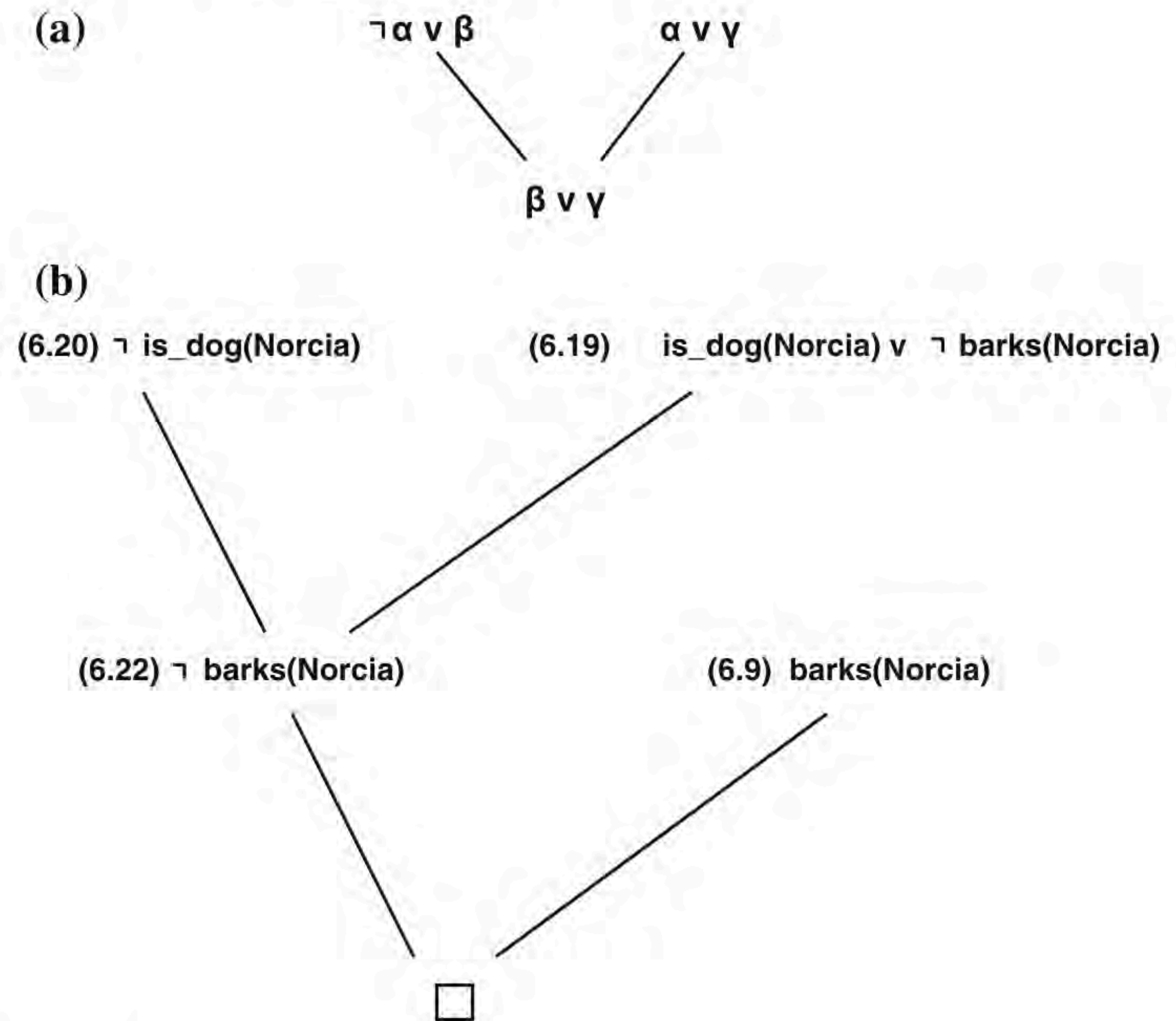


Fig. 6.2 A resolution tree

GOFAI: „Semantische Netzwerke“

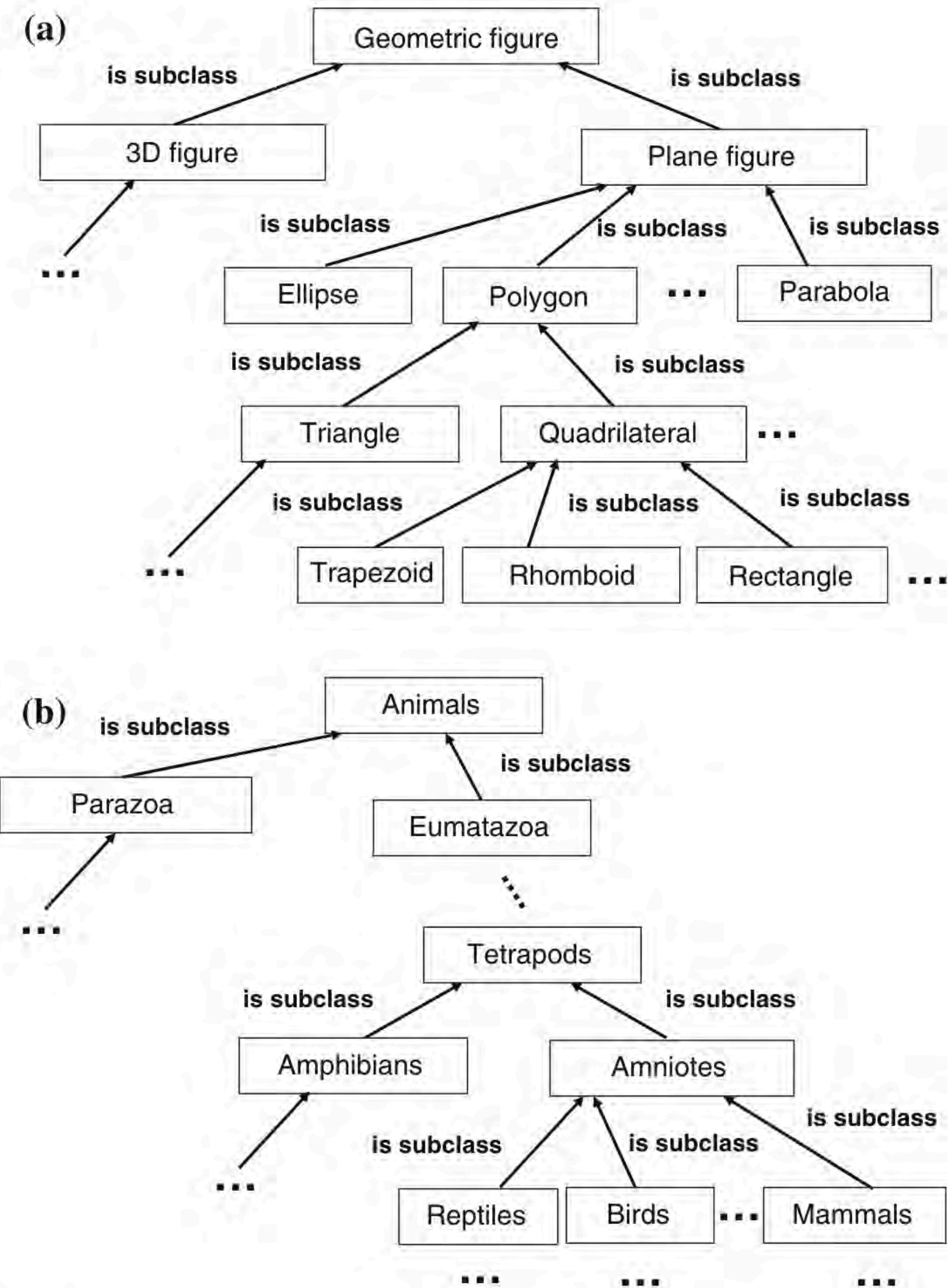


Fig. 7.1 Examples of simple semantic networks (ontologies): **a** in geometry, **b** in biology

Markov-Ketten: Zukunft als Wahrscheinlichkeitsfeld

„Future is independent of the past given the present“

Socio-Econ. Plan. Sci. Vol. 7, pp. 283–294 (1973). Pergamon Press. Printed in Great Britain

THE APPLICATION OF A MARKOV CHAIN IN EDUCATIONAL PLANNING

JAMES N. JOHNSTONE and HUGH PHILP

School of Education, Macquarie University, North Ryde, New South Wales 2113, Australia

(Received 24 September 1972)

Mathematical models can assist educators in the preparation of their educational plans and their potential in this regard is being increasingly realized. As a result, models have found application at all levels at which planning is conducted.

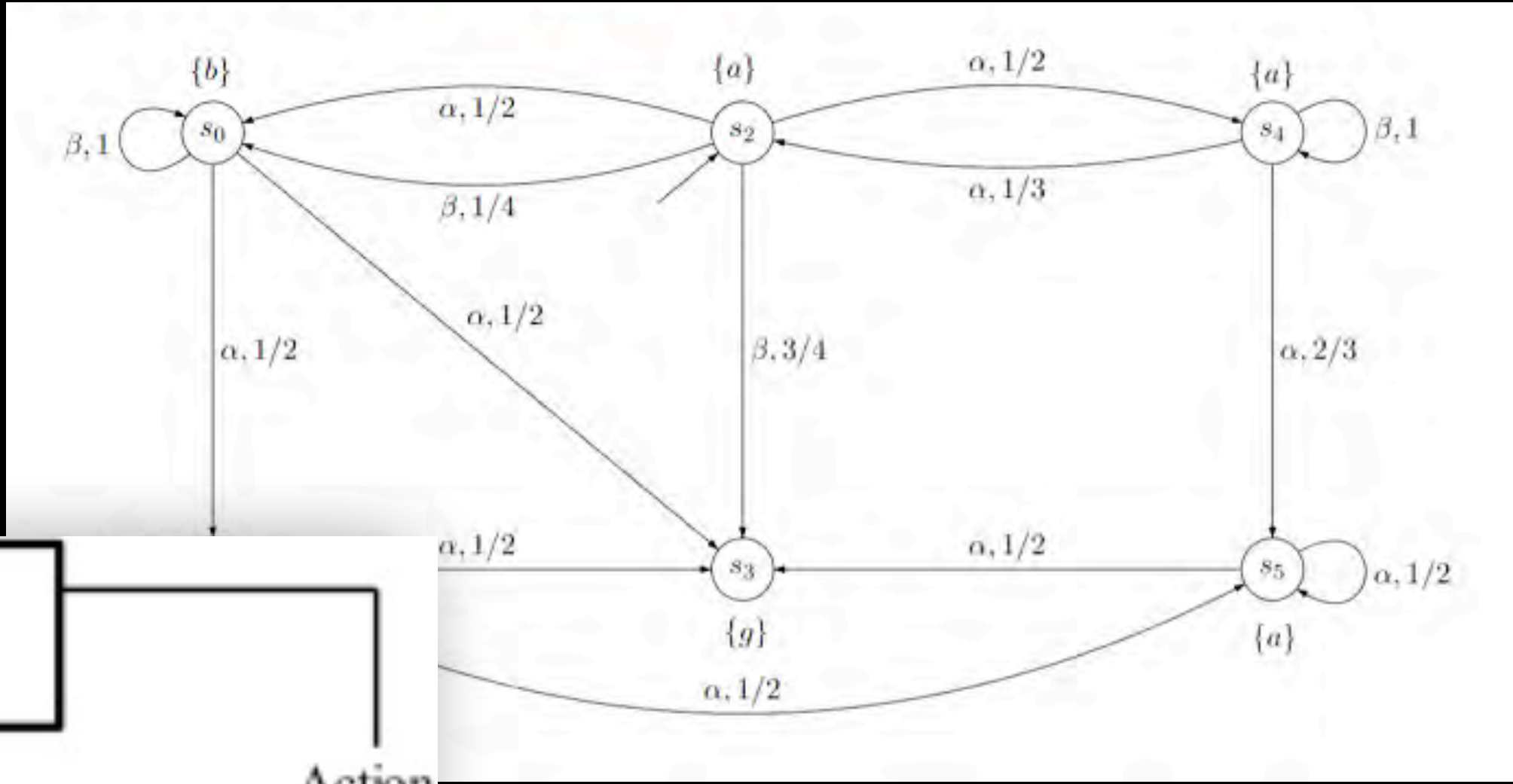
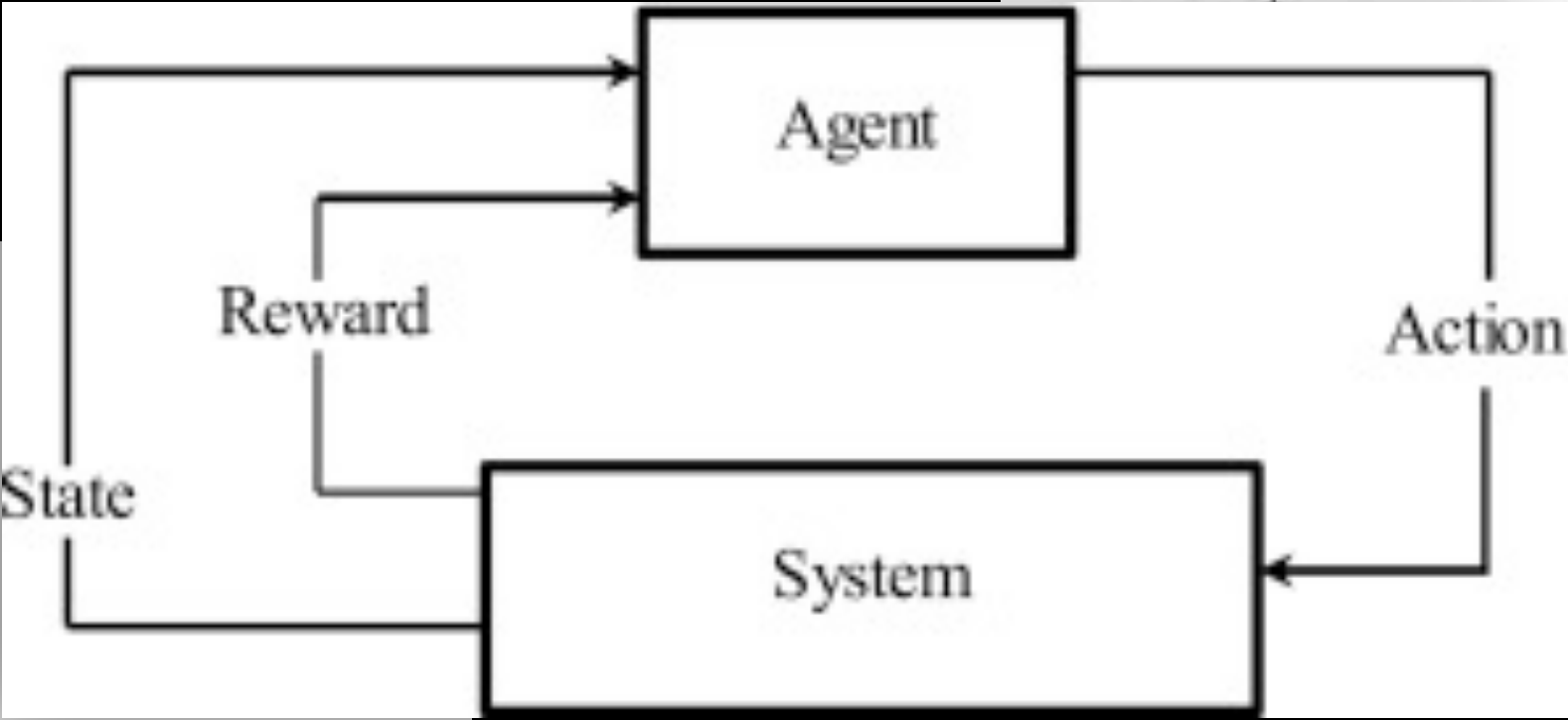
This paper examines the potential of one model—the Markov Chain—which is capable of predicting enrolments for an education system.

The model is applied to the New South Wales State Government education system between 1947 and 1961 and the projected enrolments compared to the actual enrolments in those years.

Some success is achieved but it appears as if the data rather than the model are responsible for this. The limitations of the Markov Chain approach are discussed and present research and directions listed.

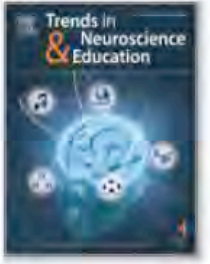
INTRODUCTION

THE APPLICATION of mathematical models to educational planning is a comparatively recent



Trends in Neuroscience and Education

Volume 5, Issue 4, December 2016, Pages 157-165



Research article

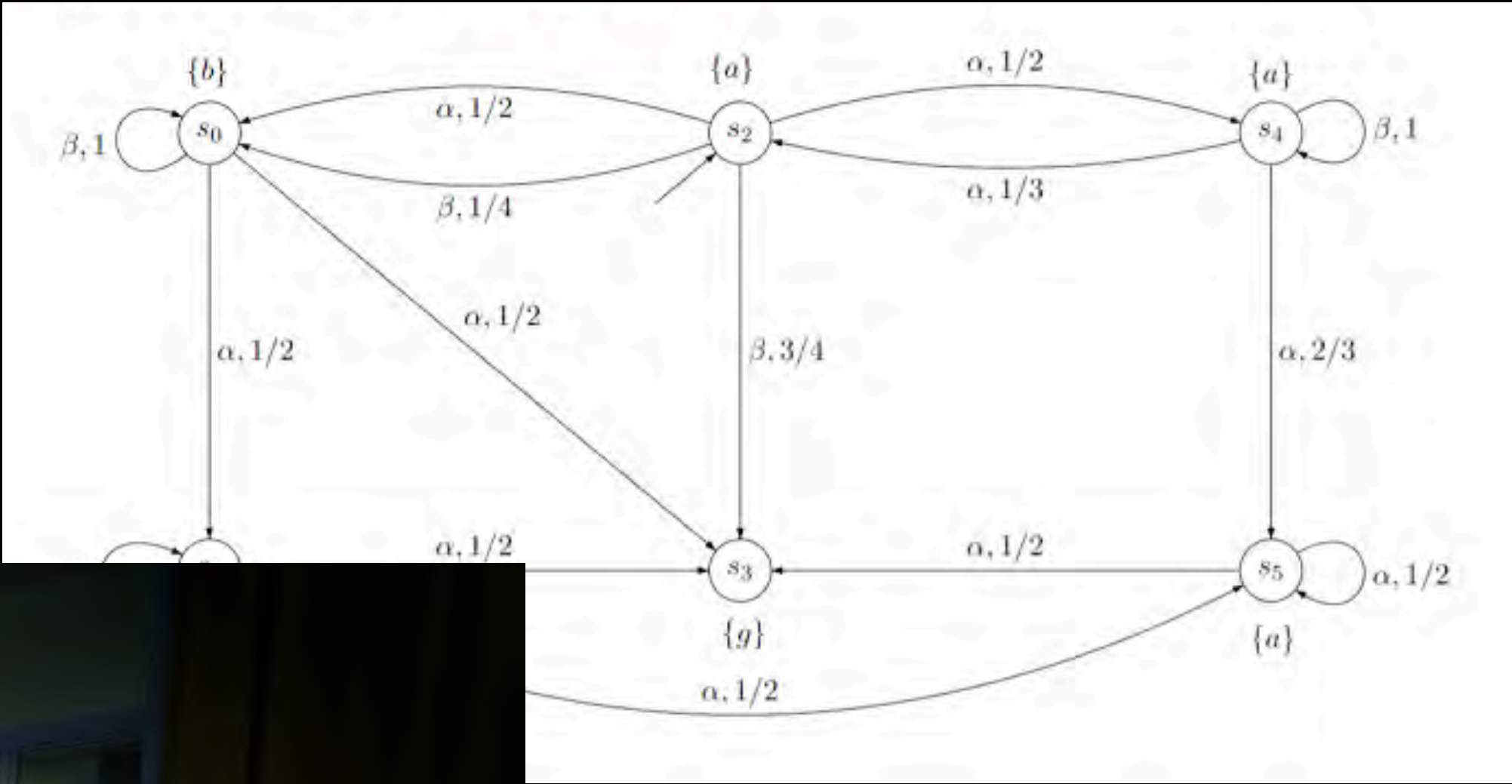
Predicting long-term outcomes of educational interventions using the evolutionary causal matrices and Markov chain based on educational neuroscience

Hyemin Han ^a, Kangwook Lee ^b, Firat Soylu ^a

Show more

+ Add to Mendeley Share Cite

Markov-Ketten in human-nonhumanen Kreativprozessen



LEONARDO

Über uns

Forschung

Projekte

Angebote

Programm

Labs

Projekte > Spirio Sessions

Spirio Sessions

Kann eine Künstliche Intelligenz auf gleichem Niveau mit einem Menschen in kreativen Schöpfungsprozessen zusammenarbeiten? Die Technische Hochschule Nürnberg und die Hochschule für Musik Nürnberg kooperieren im Forschungsprojekt „Spirio Sessions“ zu Künstlicher Kreativität.



Valentina Oefele improvisiert mit dem Markovketten-Prototypen das Spirio Sessions-Projekts

<https://youtu.be/3wDNCdrbrgE>

generative Software als algorithmischer, kultursensibler Dialogpartner und als dekoloniales Befreiungsprogramm



WAYS & MEANS

Too Many Notes: Computers, Complexity and Culture in *Voyager*

George E. Lewis

Voyager [1,2] is a nonhierarchical, interactive musical environment that privileges improvisation. In *Voyager*, improvisors engage in dialogue with a computer-driven, interactive “virtual improvising orchestra.” A computer program analyzes aspects of a human improvisor’s performance in real time, using that analysis to guide an automatic composition (or, if you will, improvisation) program that generates both complex responses to the musician’s playing and independent behavior that arises from its own internal processes.

This work, which is one of my most widely performed compositions, deals with the nature of music and, in particular, the processes by which improvising musicians produce it. These questions can encompass not only technological or music-theoretical interests but philosophical, political, cultural and social concerns as well. This is consistent with the instrumental dimension or tendency in African musical organization, or what Robert Farris Thompson [3] identifies as “songs and dances of social allusion,” one of several “ancient African organizing principles of song and dance that crossed the seas from the Old World to the New.”

Voyager’s unusual amalgamation of improvisation, indeterminacy, empathy and the logical, utterly systematic structure of the computer program is described throughout this article not only as an environment, but as a “program,” a “system” and a “composition,” in the musical sense of that term. In fact, the work can take on aspects of all of these terms simultaneously—considering the conceptual level, the process of creating the software and the real-time, real-world encounter with the work as performer or listener. Flowing across these seemingly rigid conceptual boundaries encourages both improvisors and listeners to recognize the inherent instability of such taxonomies.

Musical computer programs, like any texts, are not “objective” or “universal,” but instead represent the particular ideas of their creators. As notions about the nature and function of music become embedded into the structure of software-based musical systems and compositions, interactions with these systems tend to reveal characteristics of the community of thought and culture that produced them. Thus, it would be useful here to examine the implications of the experience of programming and performing with *Voyager* as a kind of computer music-making embodying African-American cultural practice.

Among the fair number of studies by artists/theorists who have written cogently on issues of race, gender and class in new technological media (such as Loretta Todd [4] and Cameron Bailey [5]), the ethnographic study of Institut Recherche et Coordination Acoustique/Musique (IRCAM) by the anthropologist and improviser Georgina Born [6] ap-

pears to stand practically alone in the trenchancy and thoroughness of its analysis of these issues with respect to computer music. This viewpoint contrasts markedly with Catherine M. Cameron’s [7] rather celebratory ethnography-at-a-distance of what she terms “American experimentalism,” in which the word “race” never appears, and in which her notion of a “musical class structure” is framed largely in terms of a now-moribund debate about relative privilege between Europe and America.

In contrast, Born’s explicit identification of the nearly all-male, all-white musical and cultural canon articulated not only by the French institute, but by its American equivalents, traces the outlines of the development of a post-1950s aesthetic of trans-European experimentalism. Given her so far unrefuted thesis that the overwhelming majority of computer music research and compositional activity locates itself (however unsteadily at times) within the belief systems and cultural practices of European concert music, one can easily imagine a work that, like *Voyager*, exemplifies an area of musical discourse using computers that is not viewed culturally and historically as a branch of trans-European contemporary concert music and, moreover, is not necessarily modeled as a narrative about “composition.”

THE AESTHETICS OF MULTIDOMINANCE

In an influential 1990s essay, the artist and critic Robert L. Douglas [8] sought to formalize an African-American aesthetic, synthesizing visual and musical elements of what the painter Jeff Donaldson, founder of the Africobra art movement [9], has called “Trans-African” culture. The aspect of Douglas’s theory that I wish to highlight here is the notion of “multidominant elements,” which I will henceforth call “multidominance.” According to Douglas, the aesthetics of multidominance, involving “the multiple use of colors in intense degrees, or the multiple use of textures, design patterns, or shapes” [10] are found quite routinely in musical

George E. Lewis (trombonist, composer, computer/installation artist), Department of Music, Critical Studies/Experimental Practices Area, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0526, U.S.A. E-mail: gilewis@ucsd.edu. Website: <http://www.ucsd.edu/music/lewis.html>.

generative Software als algorithmischer, kultursensibler Dialogpartner und als dekoloniales Befreiungsprogramm



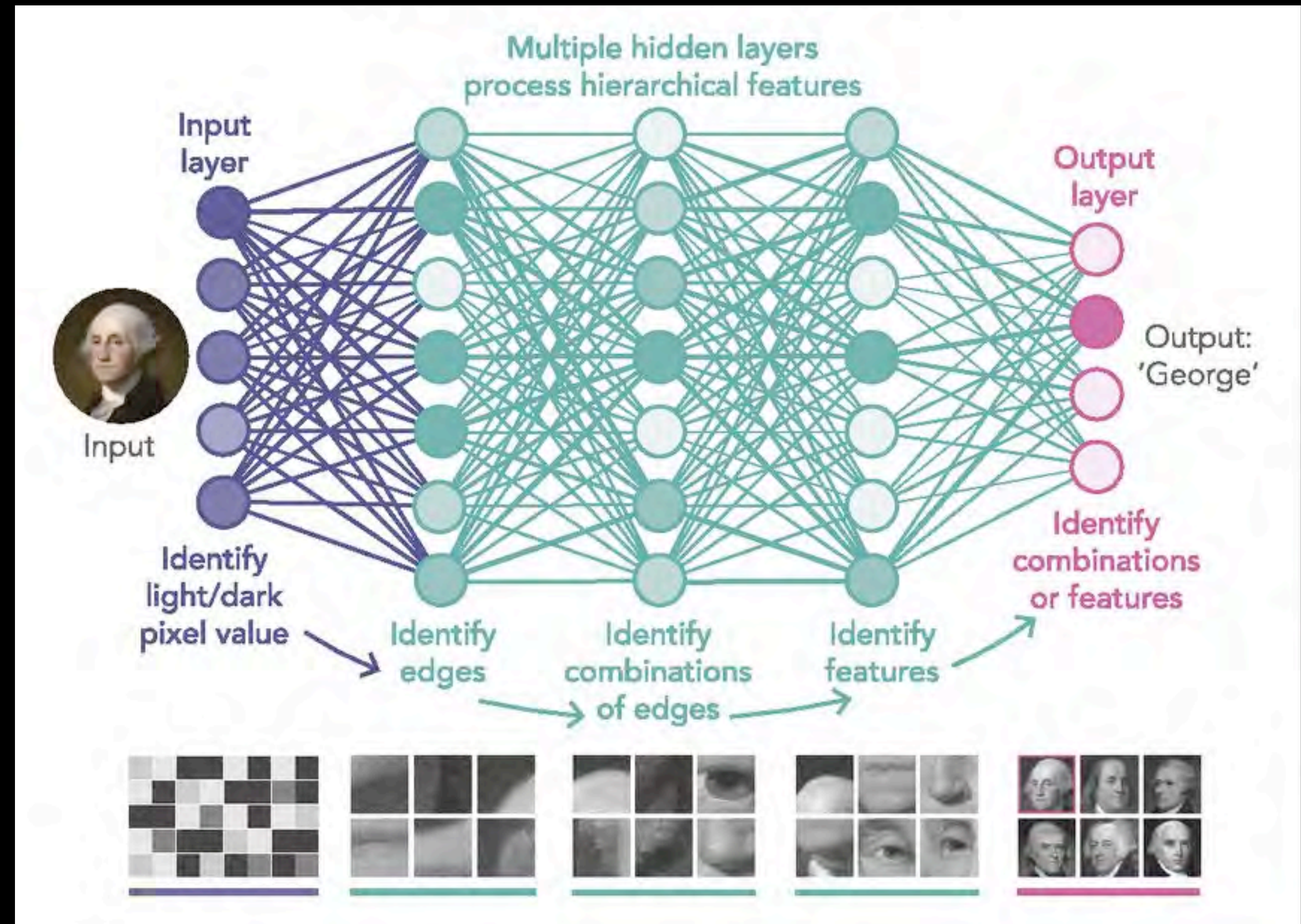
Finally, Wilson notices in African-derived music a tendency toward a high density of events in a relatively short time frame [33]. It is to be noted that the work of many important African-American improvisors—in particular Cecil Taylor, John Coltrane and Albert Ayler—exhibit a notion of extended form that involves the sustained use, often for very long periods, of extremely rapid, many-noted intensity structures. Donaldson's 1988 visual work *Jam Packed and Jelly Tight* [34] exemplifies the approach of the Africobra artists, who, according to Douglas,

used the jumpack and jelly-tight concept as a means of filling up the void, to add as much as possible to the act of creation. Africobra members accept these concepts as an African axiom: that to add to life is to ensure that there is more to share [35].

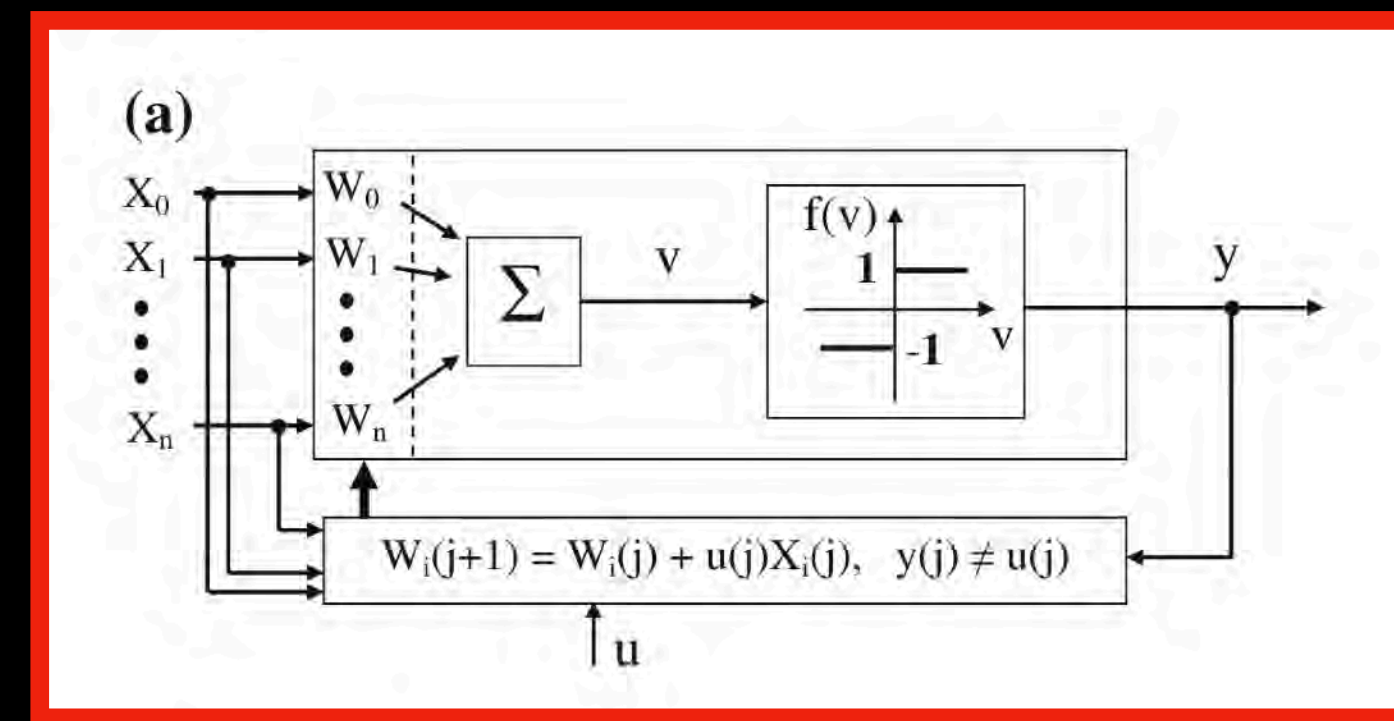
The *Voyager* program often combines dense, rapid accretions of sonic information with sudden changes of mood, tempo and orchestration, eschewing the slowly moving timbral narratives characteristic of much institutionally based com-

Non-Humane Mimesis:
Die Quantifizierung von
Gestalt- und
Ähnlichkeitsverhältnissen als
Basis von KI

Neuronales „feed forward“ Netzwerk



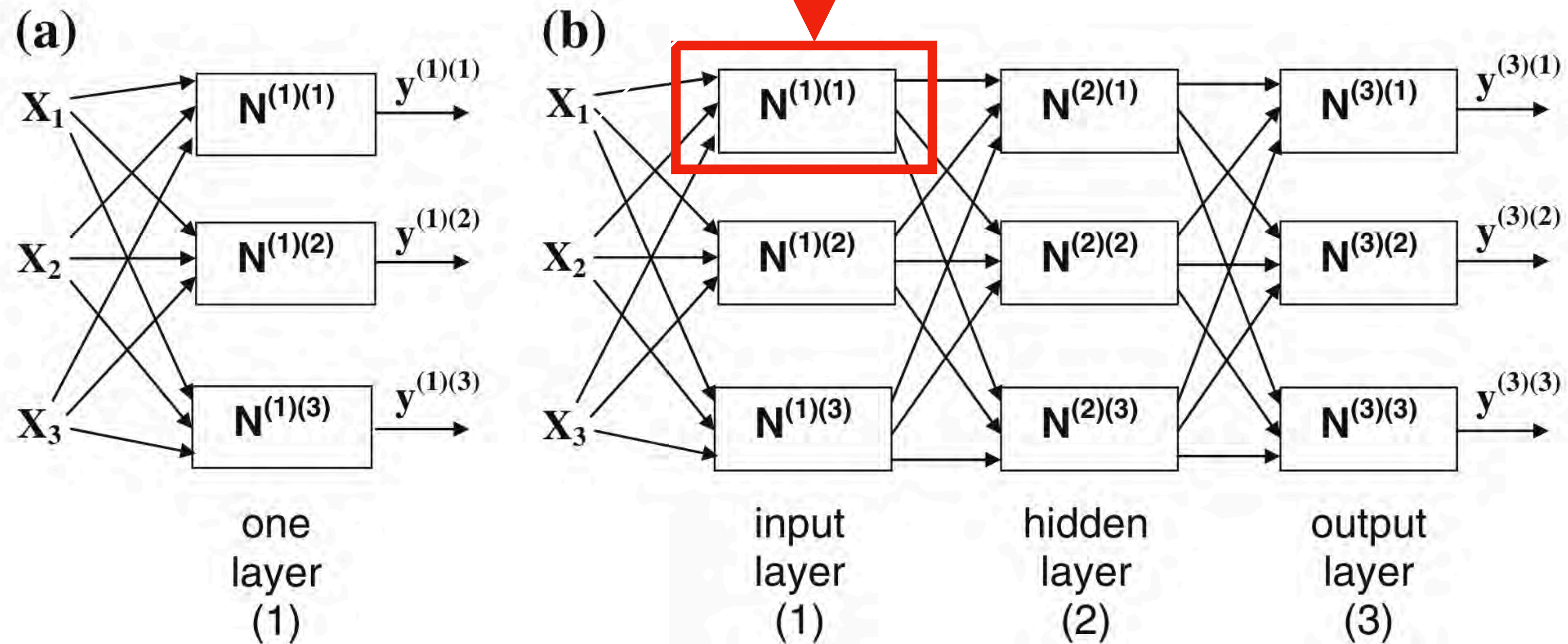
Künstliches Neuron (hier: Perceptron)



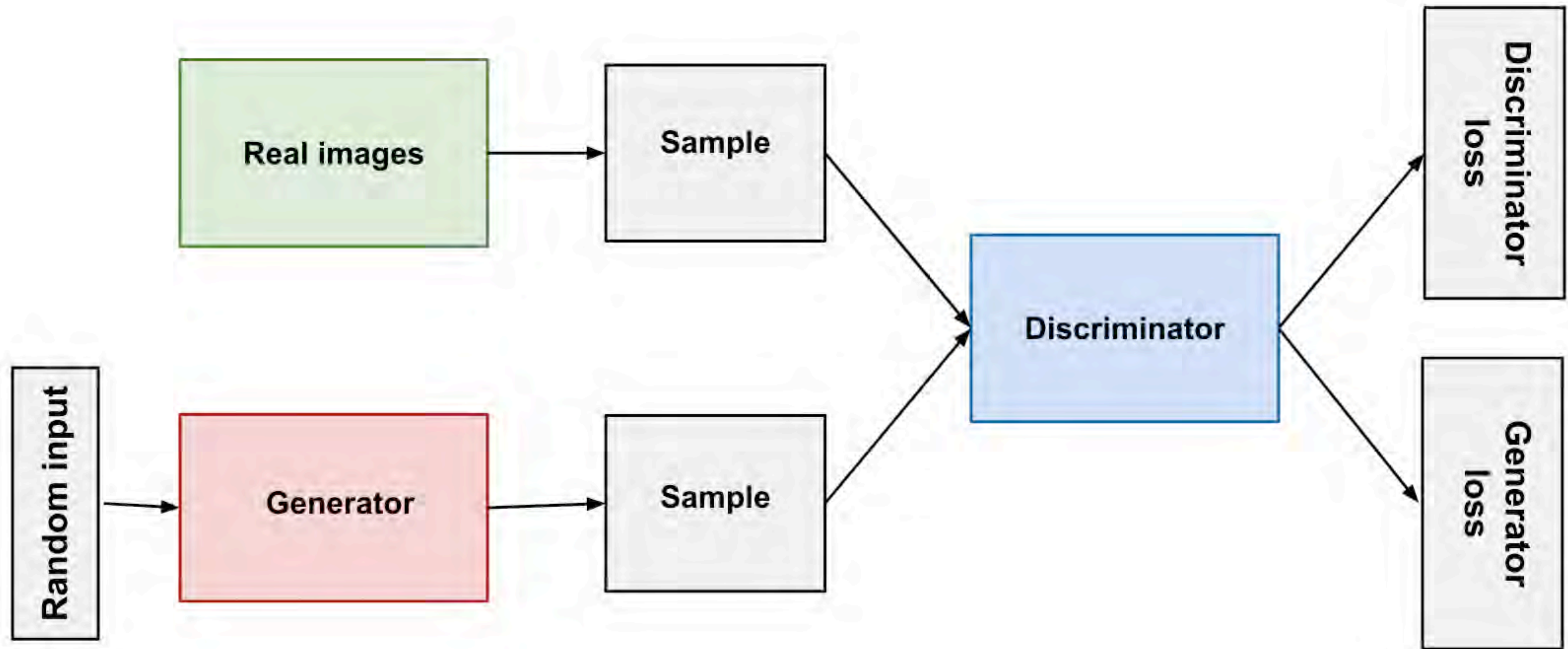
Neuronales „feed forward“ Netzwerk

168

11 Neural Networks



„Generative Adversarial Networks“ (GAN-Framework)



CHRISTIE'S
AUCTIONS
RESULTS
LOCATIONS
DEPARTMENTS
STORIES



SALE 16388
Prints & Multiples
New York | 23 - 25 October 2018



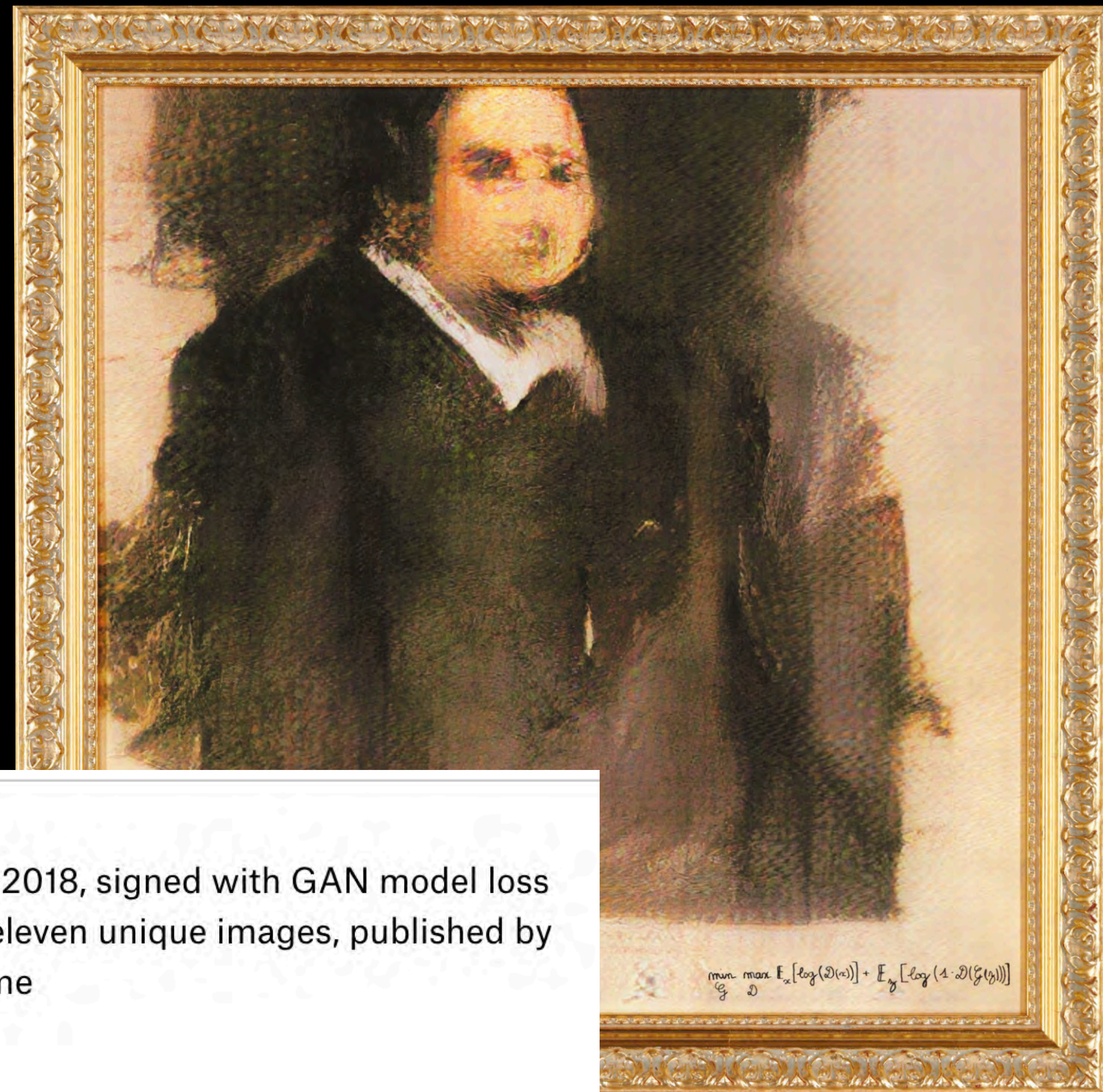
LOT 363
Edmond de Belamy, from La

Price realised ⓘ
USD 432,500

Estimate ⓘ
USD 7,000 - USD 10,000

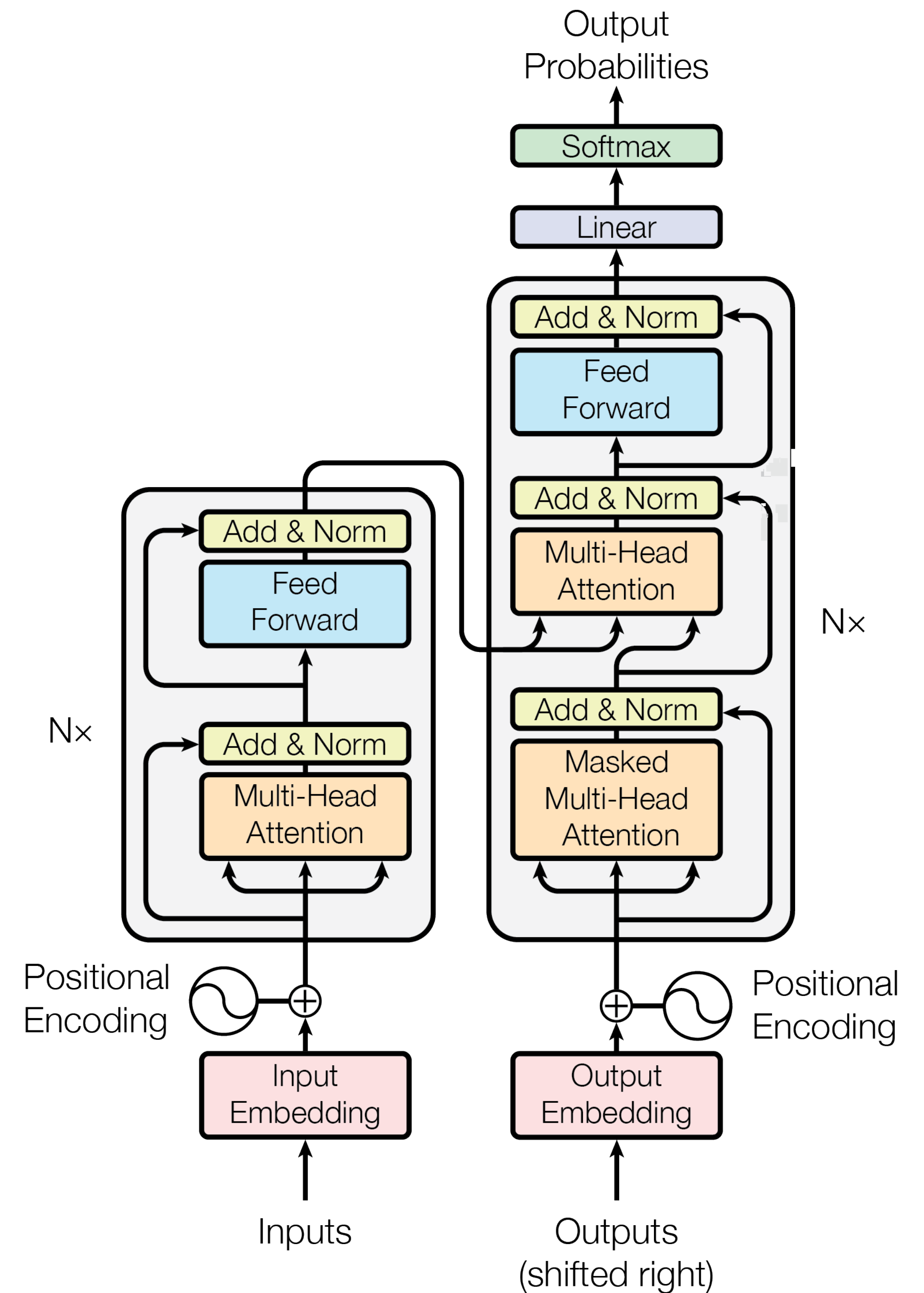
Follow lot

+ Add to Interests



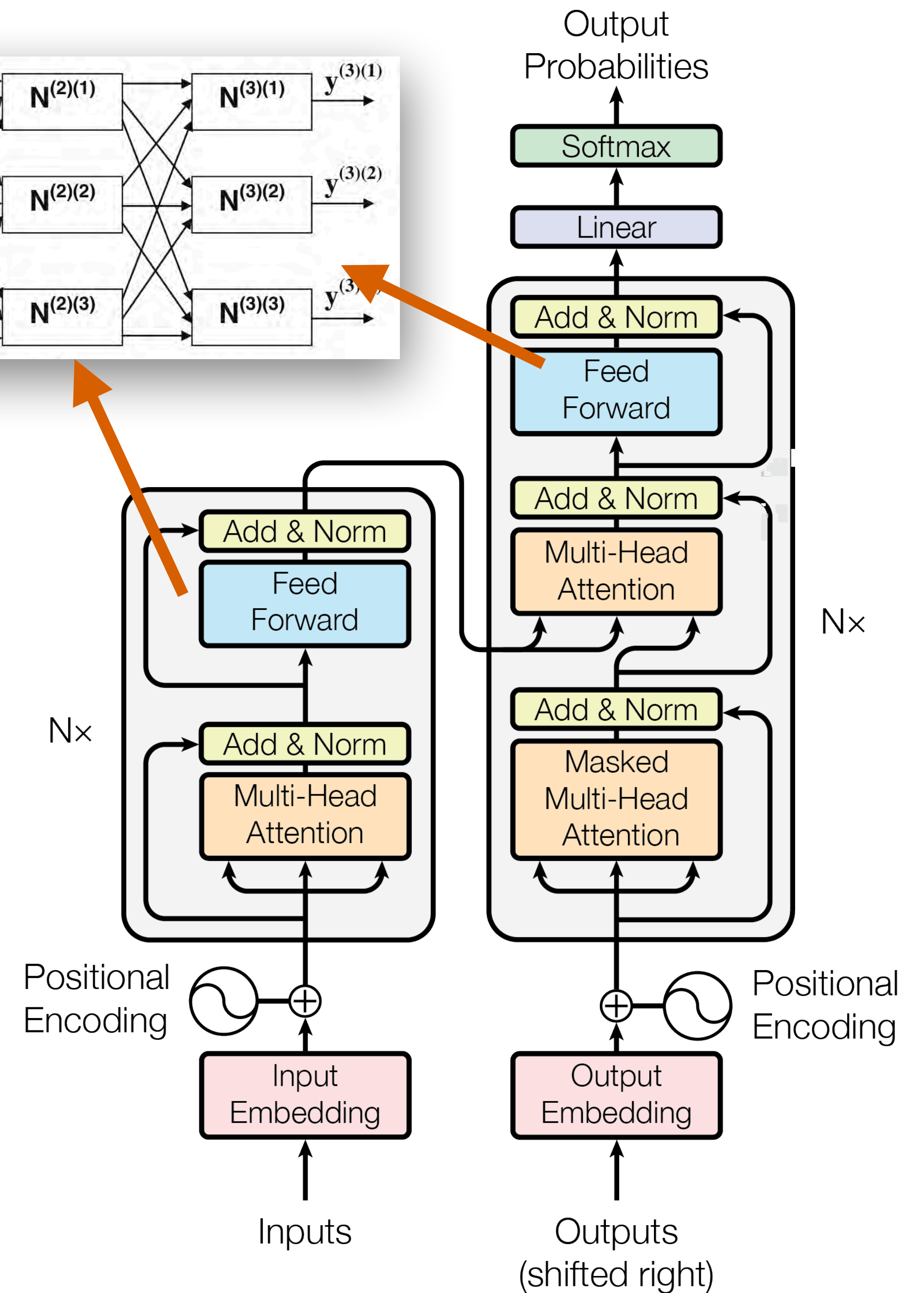
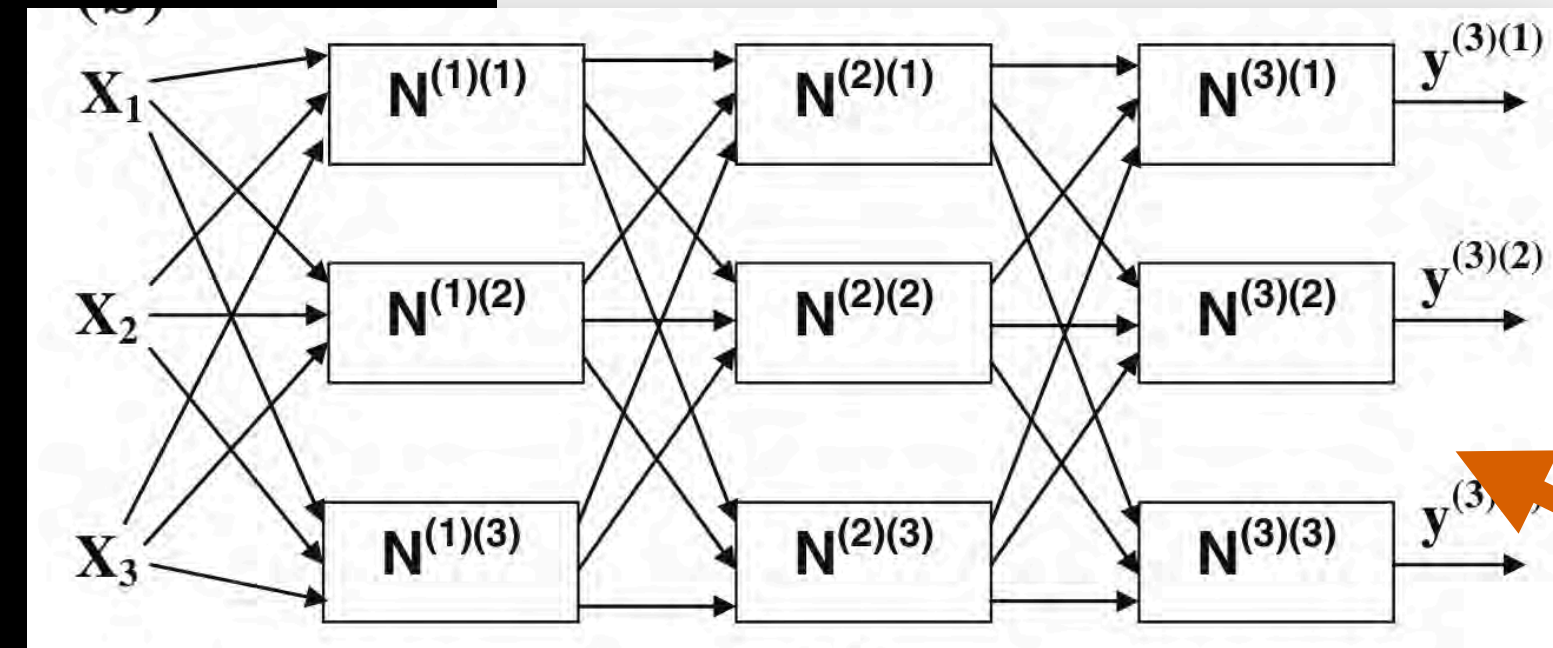
Edmond de Belamy, from *La Famille de Belamy*
generative Adversarial Network print, on canvas, 2018, signed with GAN model loss
function in ink by the publisher, from a series of eleven unique images, published by
Obvious Art, Paris, with original gilded wood frame
S. 27 ½ x 27 ½ in (700 x 700 mm.)

„Transformer Neural Networks“ (TNN): Grundlage von ChatGPT (z.B. BERT; GPT-4)

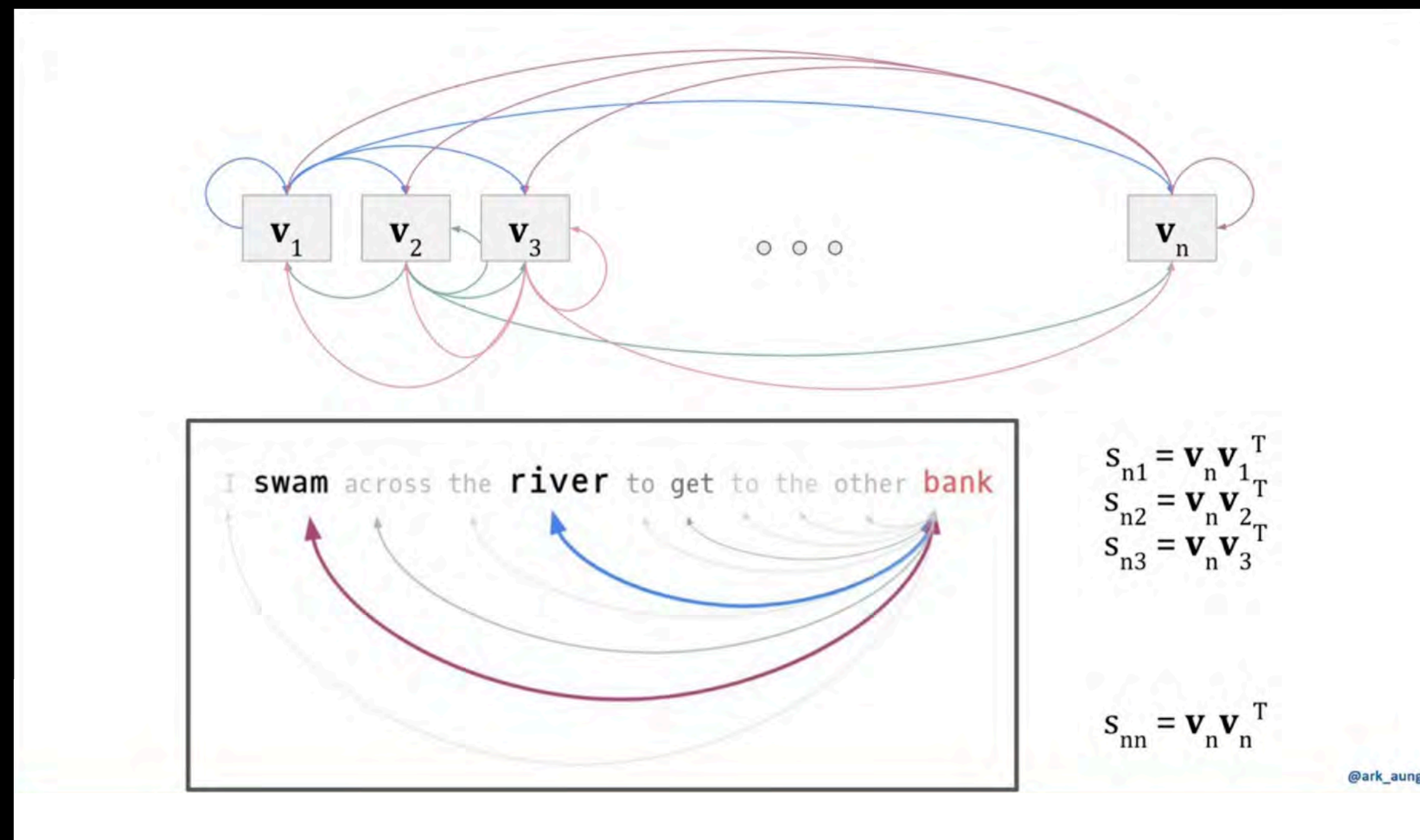


Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need (arXiv:1706.03762). arXiv. <https://doi.org/10.48550/arXiv.1706.03762>

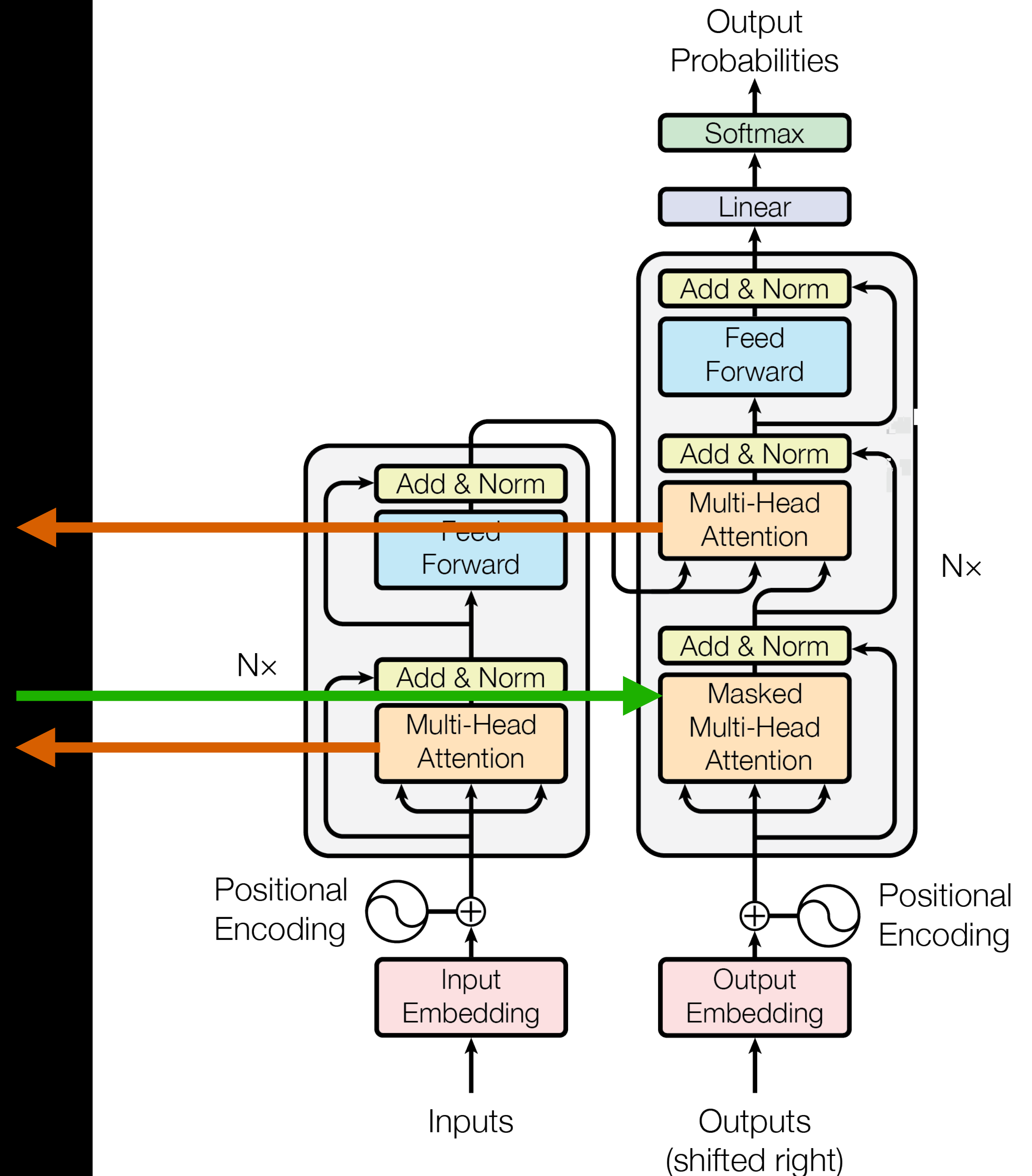
„Transformer Neural
Networks“ (TNN):
Grundlage von ChatGPT
(z.B. BERT; GPT-4)

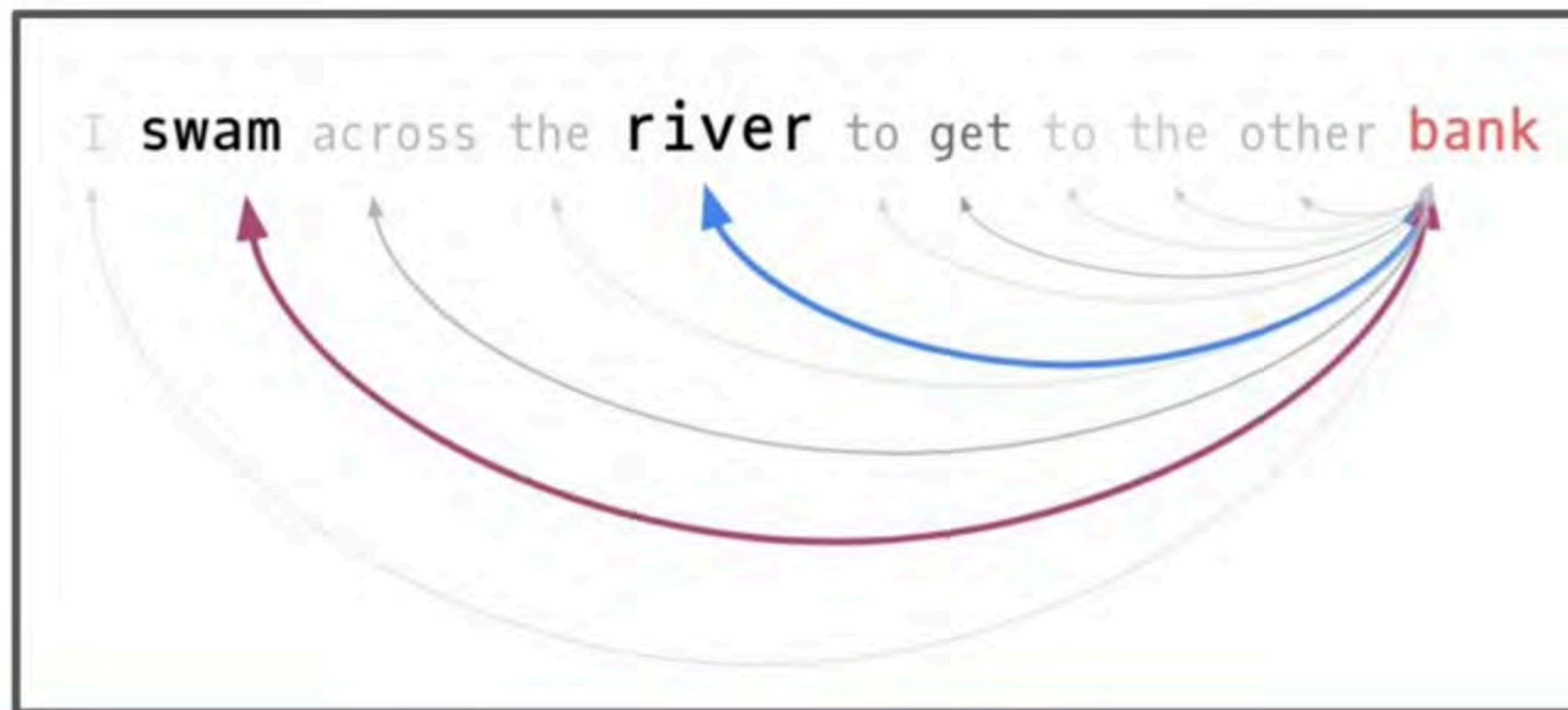
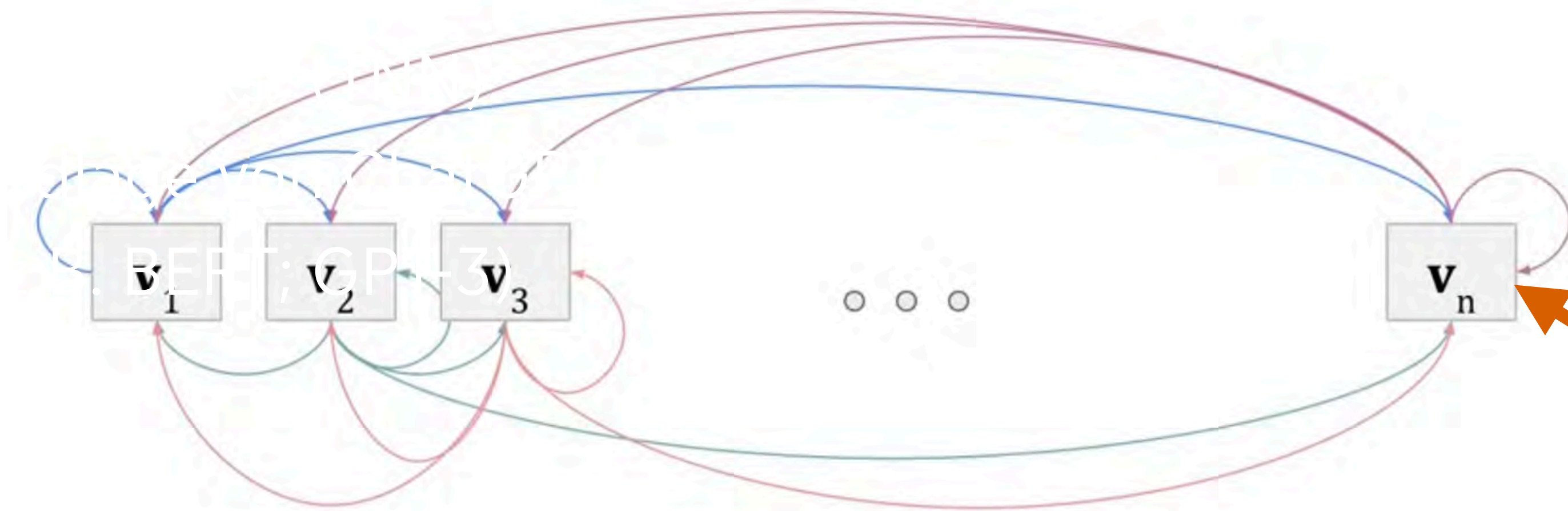


„Transformer Neural Networks“ (TNN): Grundlage von ChatGPT (z.B. BERT; GPT-4)



Aung, A. M. (Director). (2020, Oktober 17). Intuition Behind Self-Attention Mechanism in Transformer Networks.
<https://www.youtube.com/watch?v=g2BRluln4uc>





$$S_{n1} = \mathbf{v}_n \mathbf{v}_1^T$$

$$S_{n2} = \mathbf{v}_n \mathbf{v}_2^T$$

$$S_{n3} = \mathbf{v}_n \mathbf{v}_3^T$$

$$S_{nn} = \mathbf{v}_n \mathbf{v}_n^T$$

@ark_aung

Add & Norm

Multi-Head Attention

Jede Einheit (Wort) wird als **Vektor** berechnet, der Beziehungen zu anderen Einheiten (Wörtern) ausdrückt.

Embedding Projector



DATA

5 tensors found

Word2Vec 10K

Label by **word** Color by **No color map**

Edit by
word ▼ Tag selection as

Load Publish Download Label

☒ Sphereize data [?](#)

Checkpoint: Demo datasets

```
Metadata: oss_data/word2vec_10000_200d_
          labels.tsv
```

UMAP T-SNE **PCA** CUSTOM

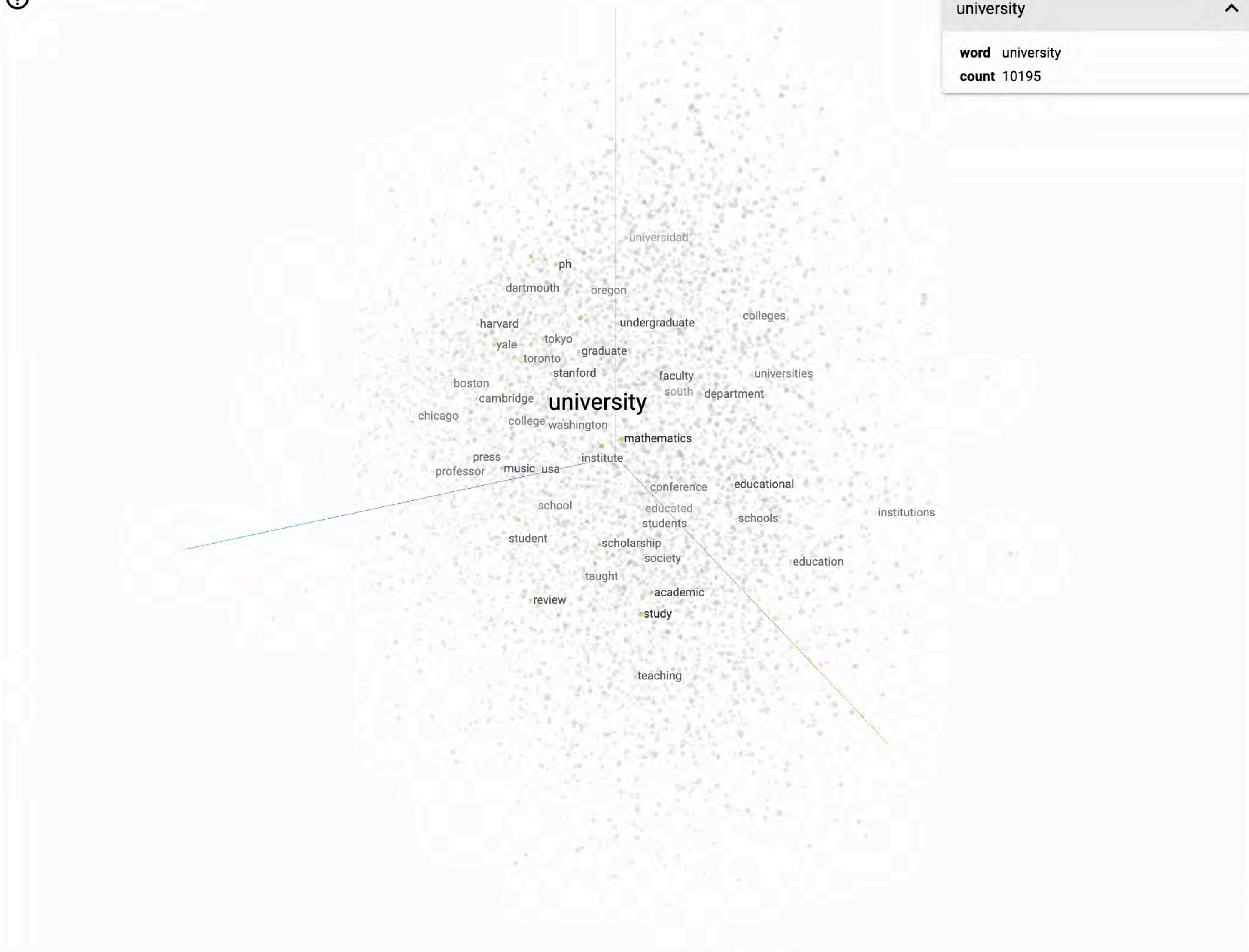
X Y

Component #3  

PCA is approximate. ?

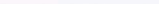
Total variance described: 8.5%.

 | Points: 10000 | Dimension: 200 | Selected 101 points



Show All Data Isolate 101 points Clear selection

Search by 

neighbors ?  100 

distance COSINE EUCLIDEAN

Nearest points in the original space:

college	0.355
school	0.403
harvard	0.453
institute	0.455
universities	0.470
professor	0.486
graduate	0.489
cambridge	0.499
stanford	0.506
oxford	0.520
yale	0.537
students	0.565
illinois	0.572
columbia	0.572
faculty	0.582
press	0.584
mit	0.585
princeton	0.585
undergraduate	0.597
education	0.602
academic	0.604
student	0.609
berkeley	0.609

BOOKMARKS (0) ? ^

Word Embeddings

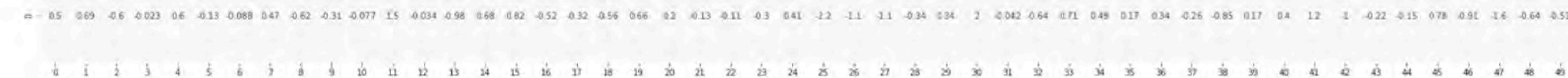
“The gift of words is the gift of deception and illusion” ~Children of Dune

With this understanding, we can proceed to look at trained word-vector examples (also called word embeddings) and start looking at some of their interesting properties.

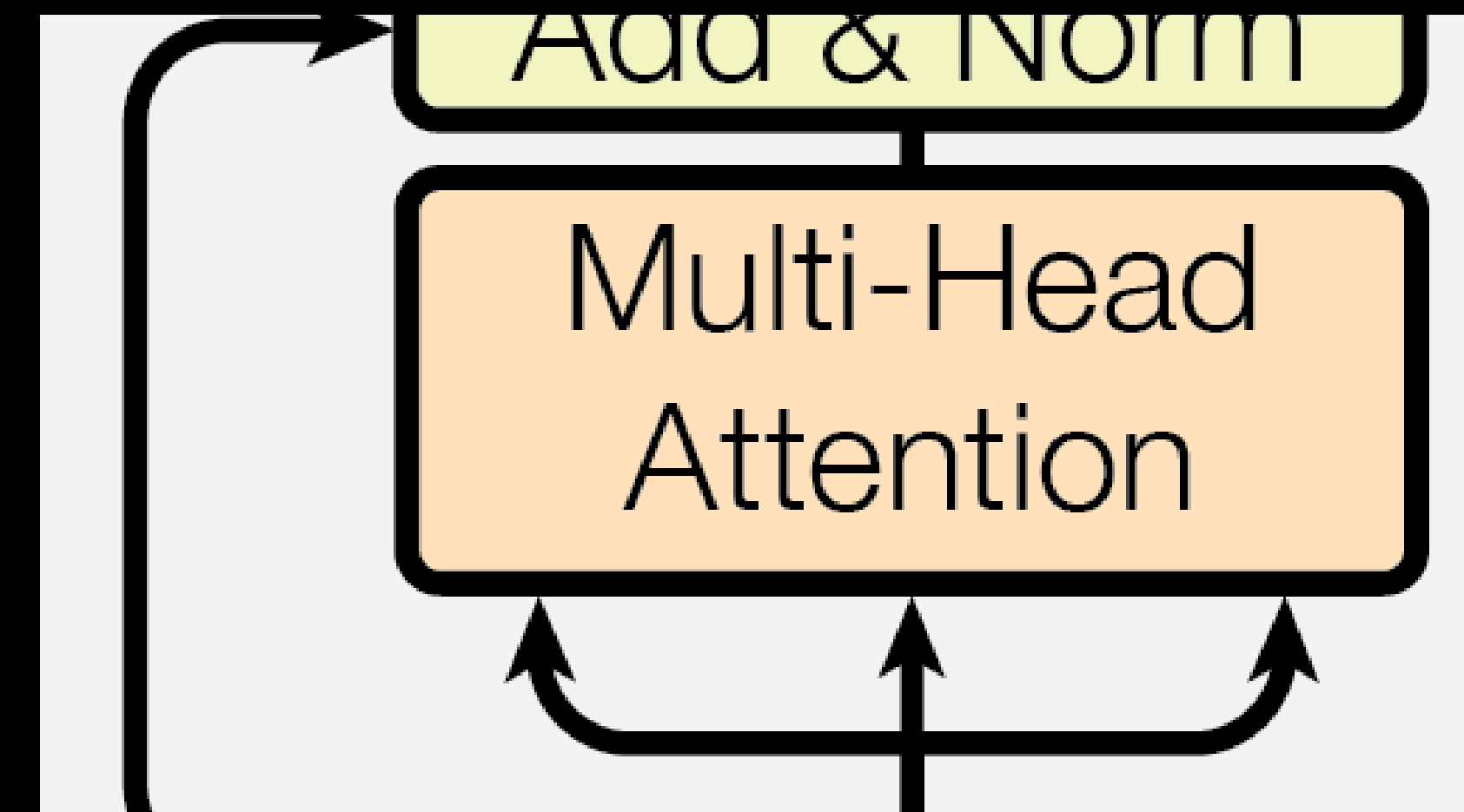
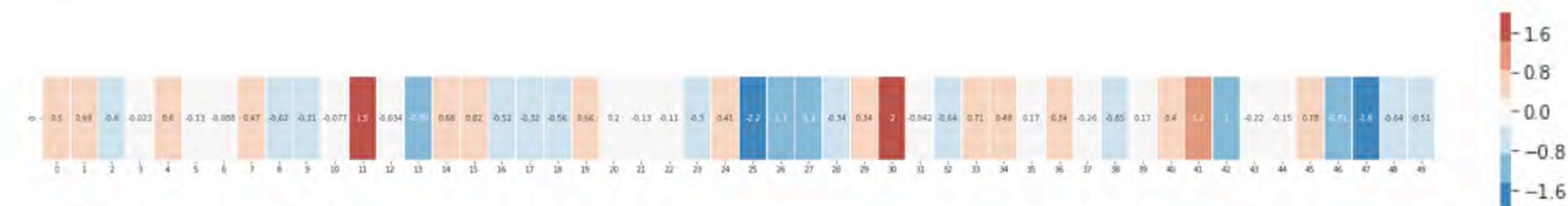
This is a word embedding for the word “king” (GloVe vector trained on Wikipedia):

[0.50451 , 0.68607 , -0.59517 , -0.022801, 0.60046 , -0.13498 , -0.08813 , 0.47377 , -0.61798 , -0.31012 ,
-0.076666, 1.493 , -0.034189, -0.98173 , 0.68229 , 0.81722 , -0.51874 , -0.31503 , -0.55809 , 0.66421 , 0.1961
, -0.13495 , -0.11476 , -0.30344 , 0.41177 , -2.223 , -1.0756 , -1.0783 , -0.34354 , 0.33505 , 1.9927 ,
-0.04234 , -0.64319 , 0.71125 , 0.49159 , 0.16754 , 0.34344 , -0.25663 , -0.8523 , 0.1661 , 0.40102 , 1.1685 ,
-1.0137 , -0.21585 , -0.15155 , 0.78321 , -0.91241 , -1.6106 , -0.64426 , -0.51042]

It’s a list of 50 numbers. We can’t tell much by looking at the values. But let’s visualize it a bit so we can compare it
other word vectors. Let’s put all these numbers in one row:



Let’s color code the cells based on their values (red if they’re close to 2, white if they’re close to 0, blue if they’re close
to -2):



We'll proceed by ignoring the numbers and only looking at the colors to indicate the values of the cells. Let's now contrast "King" against other words:

"king"



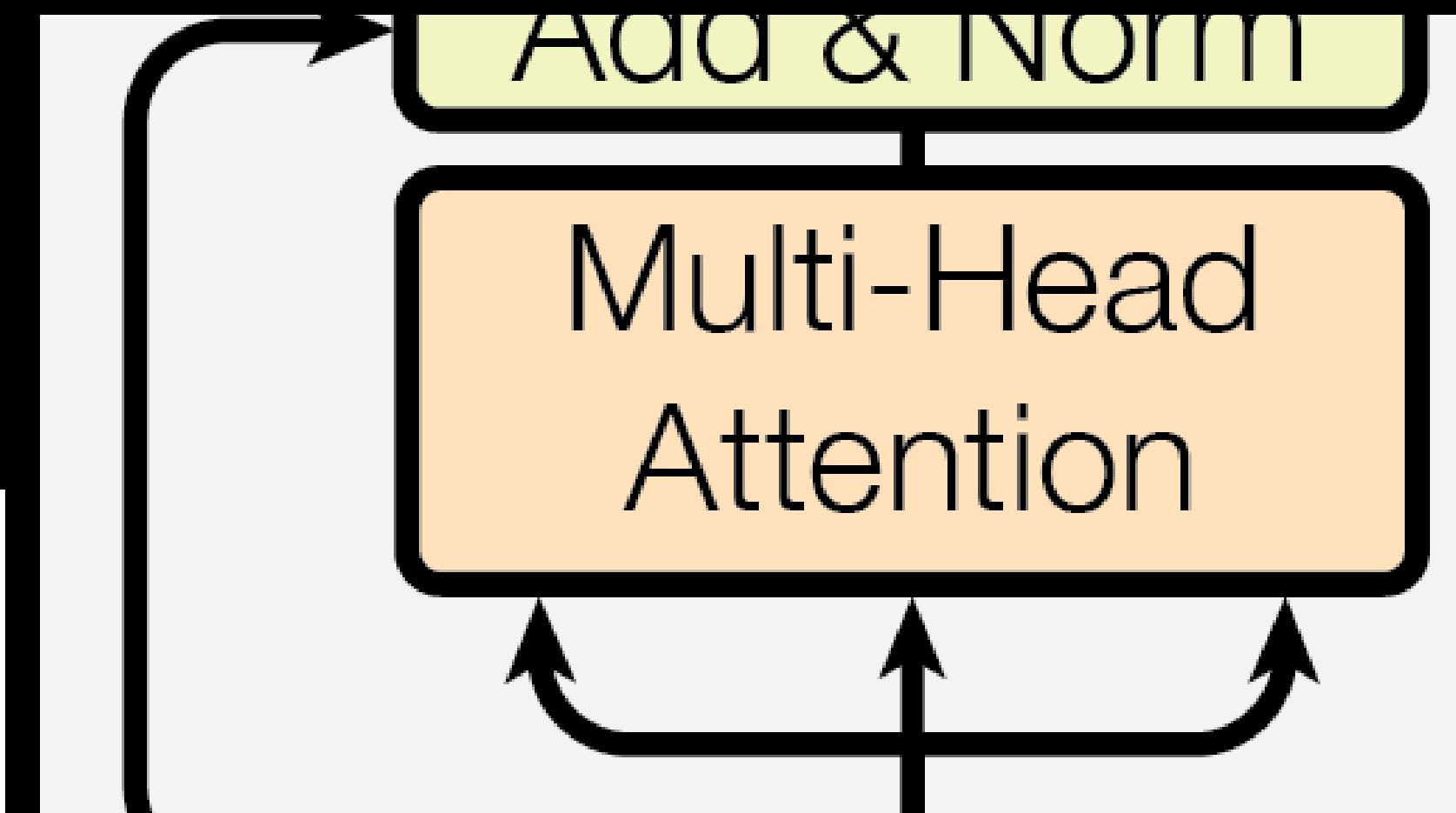
"Man"



"Woman"



See how "Man" and "Woman" are much more similar to each other than either of them is to "king"? This tells you something. These vector representations capture quite a bit of the information/meaning/associations of these words.



queen

woman

girl

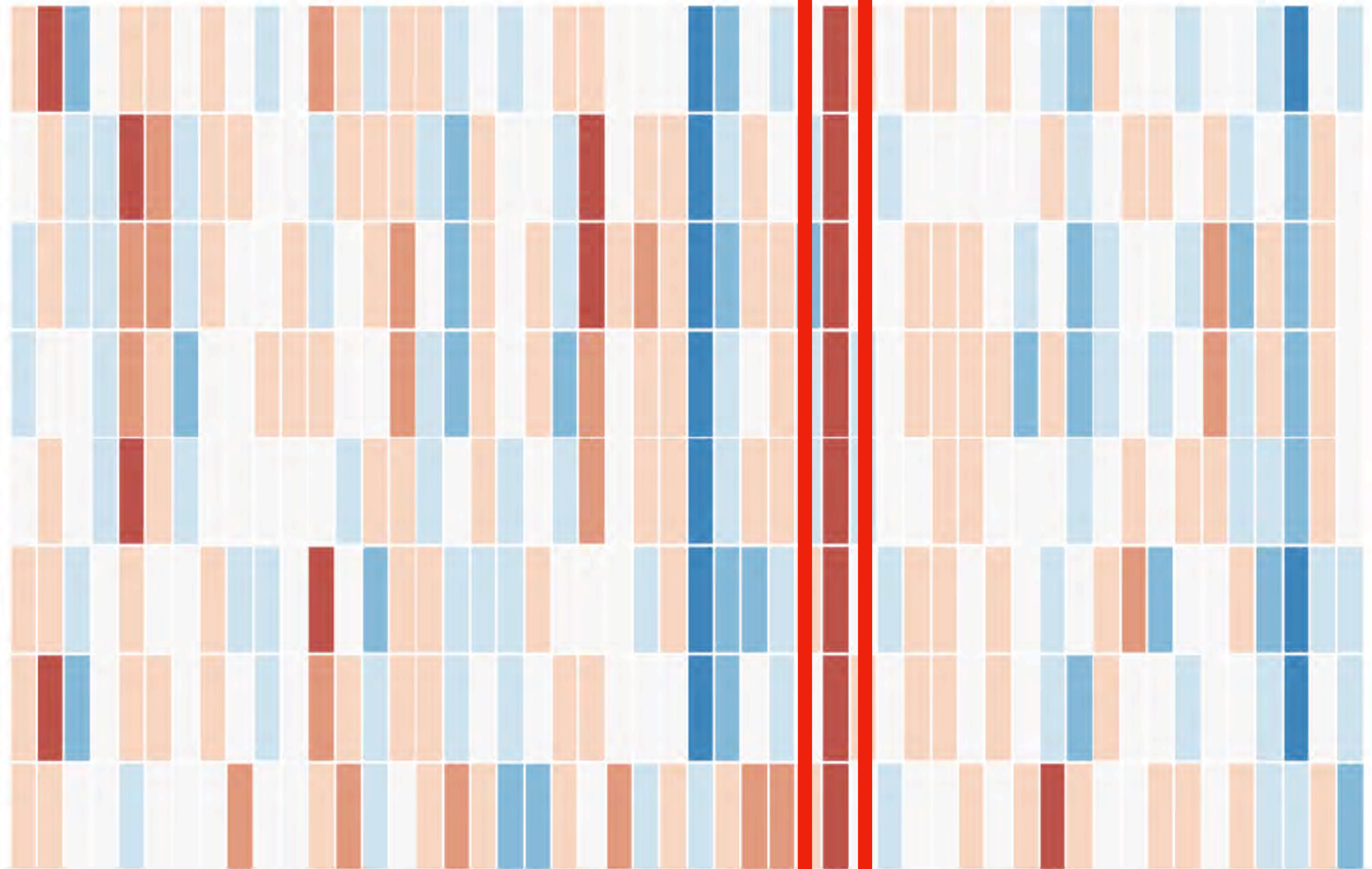
boy

man

king

queen

water



queen

woman

girl

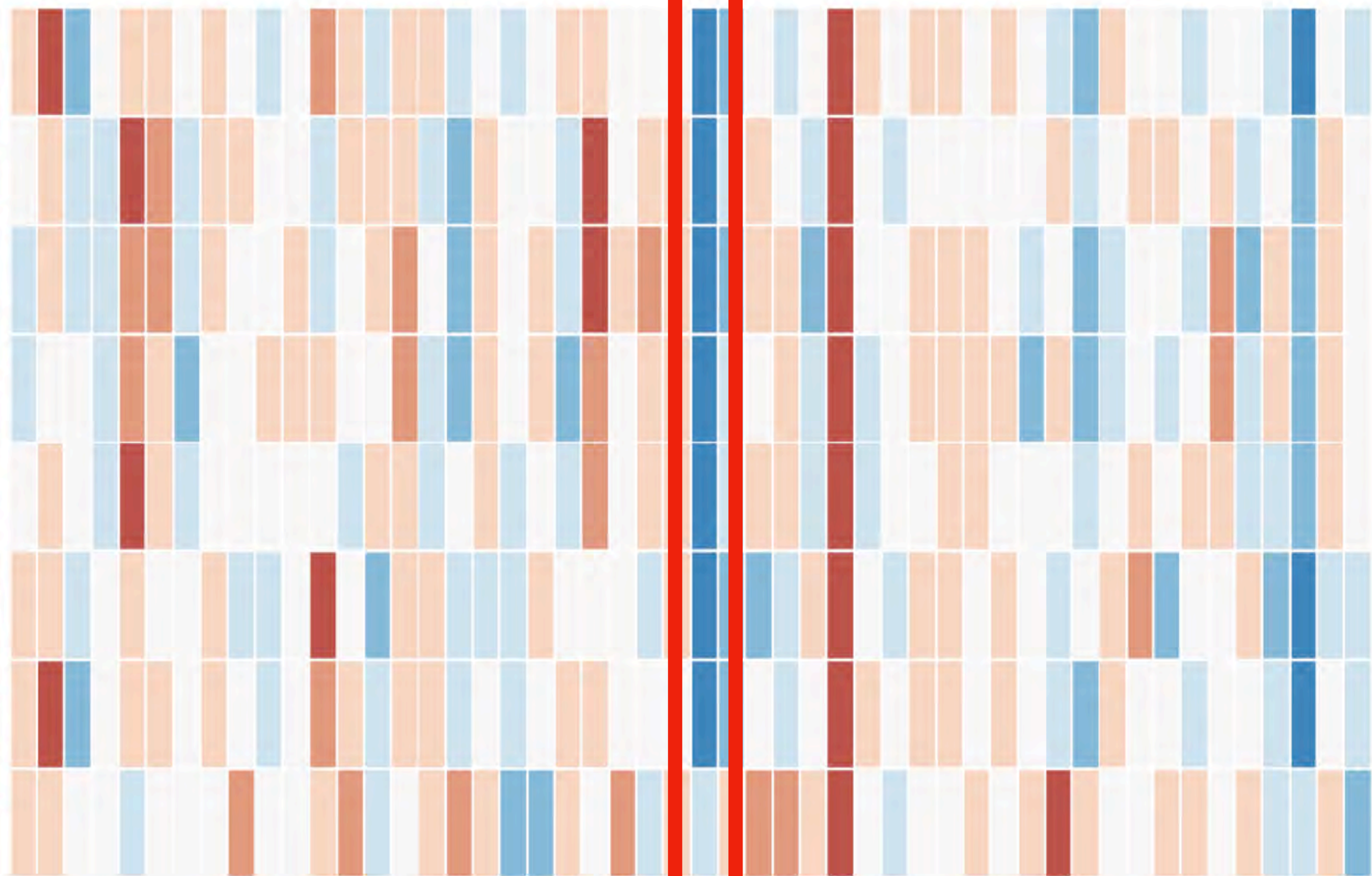
boy

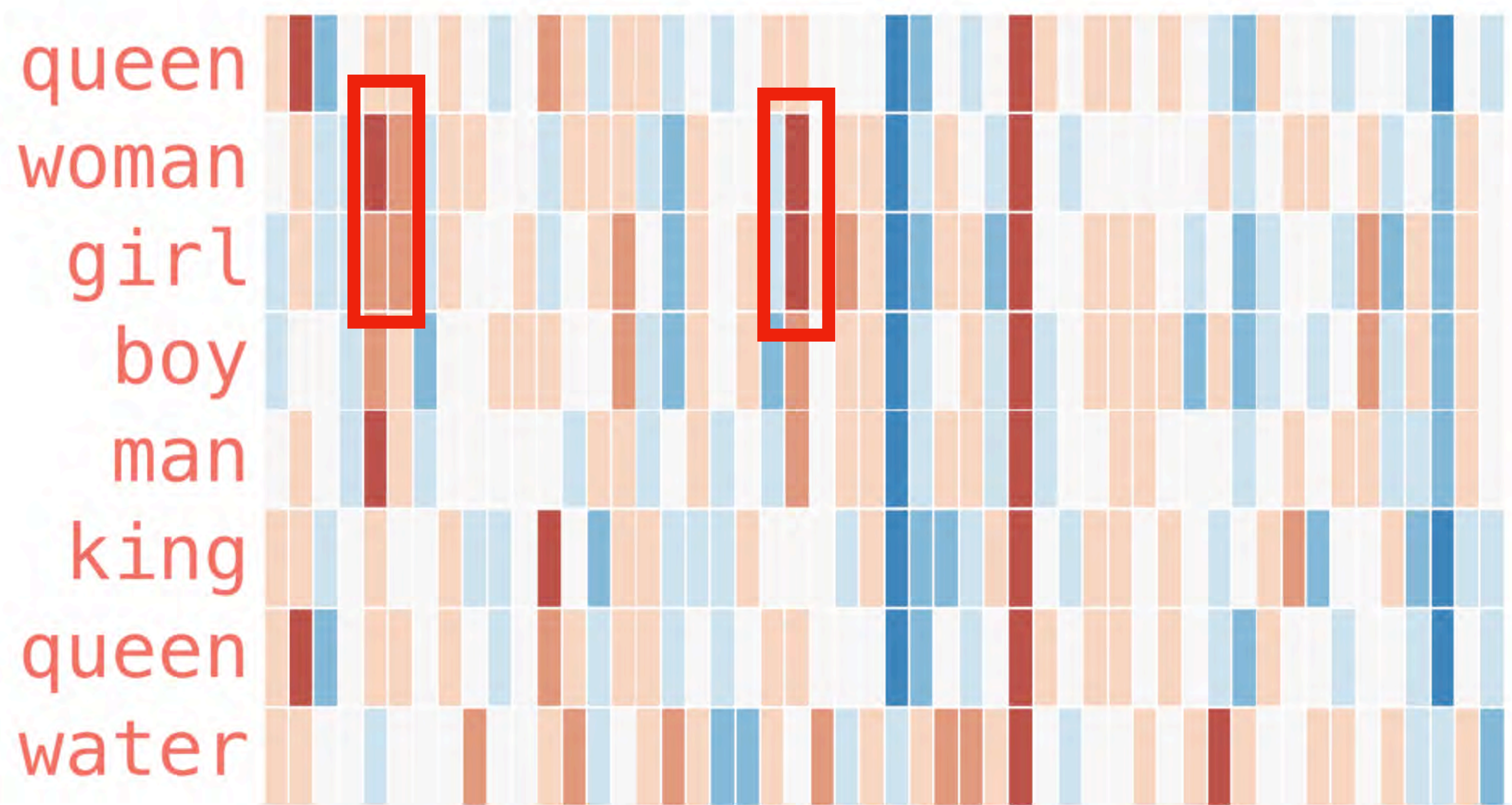
man

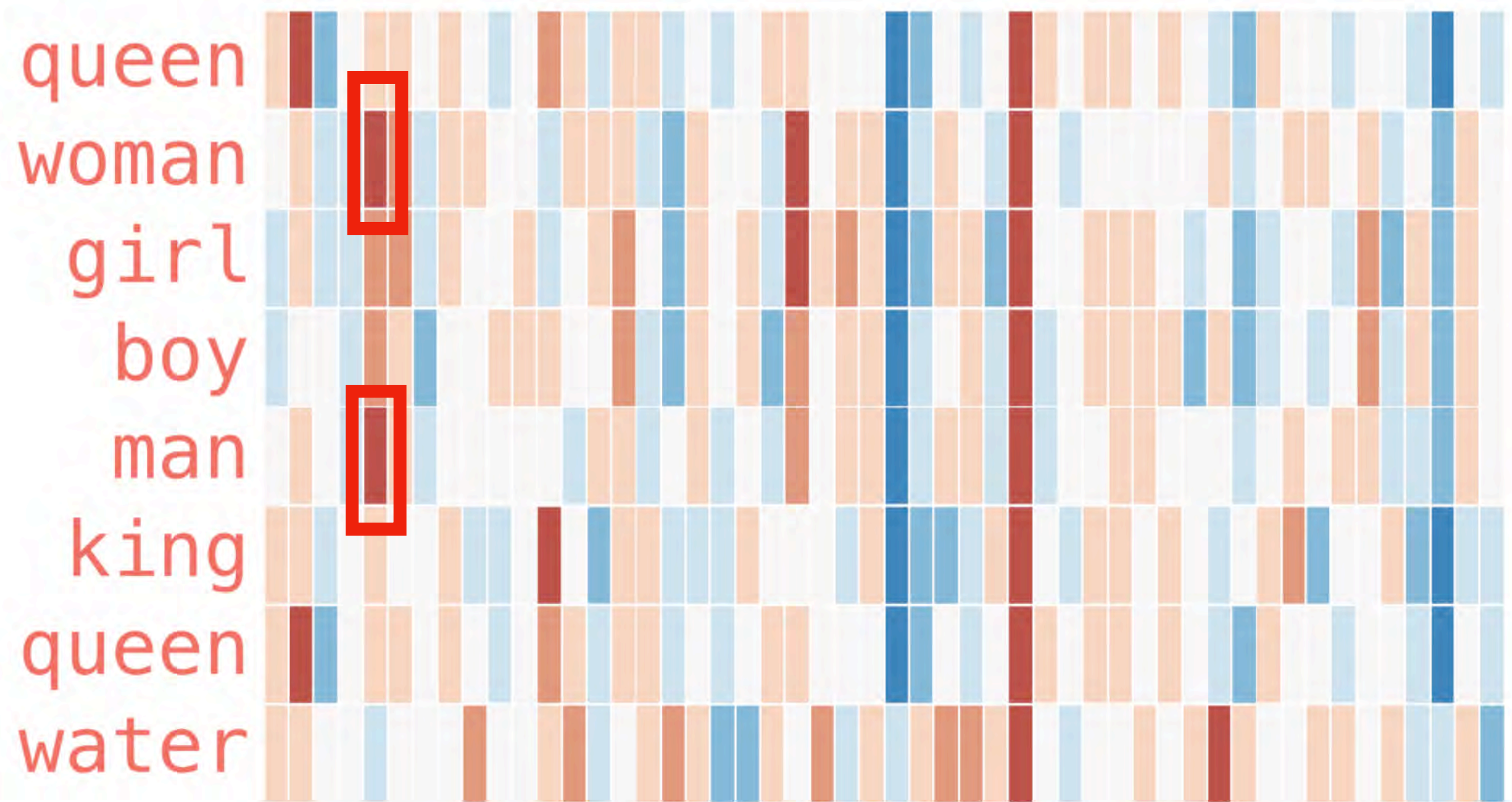
king

queen

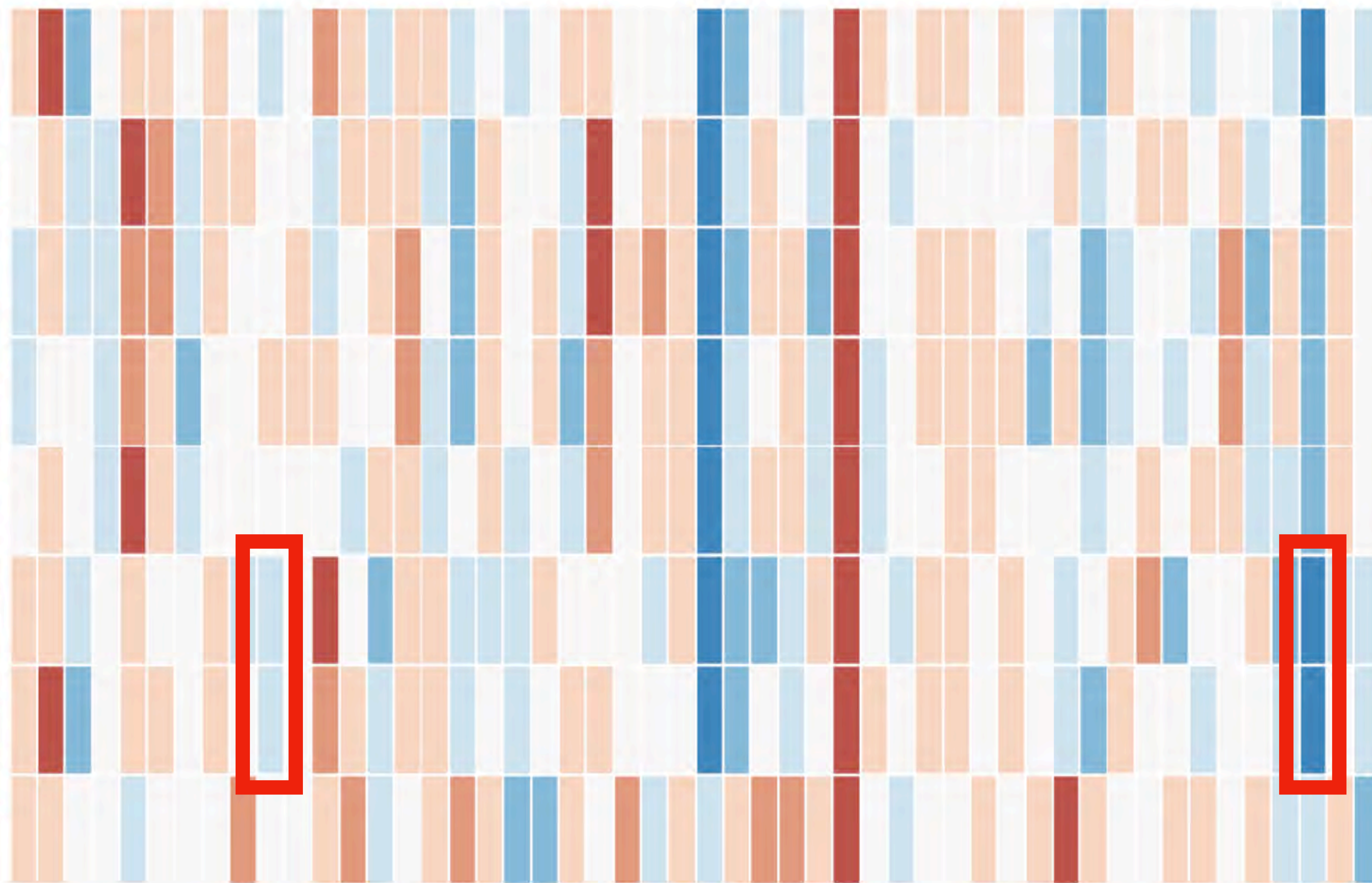
water







queen
woman
girl
boy
man
king
queen
water

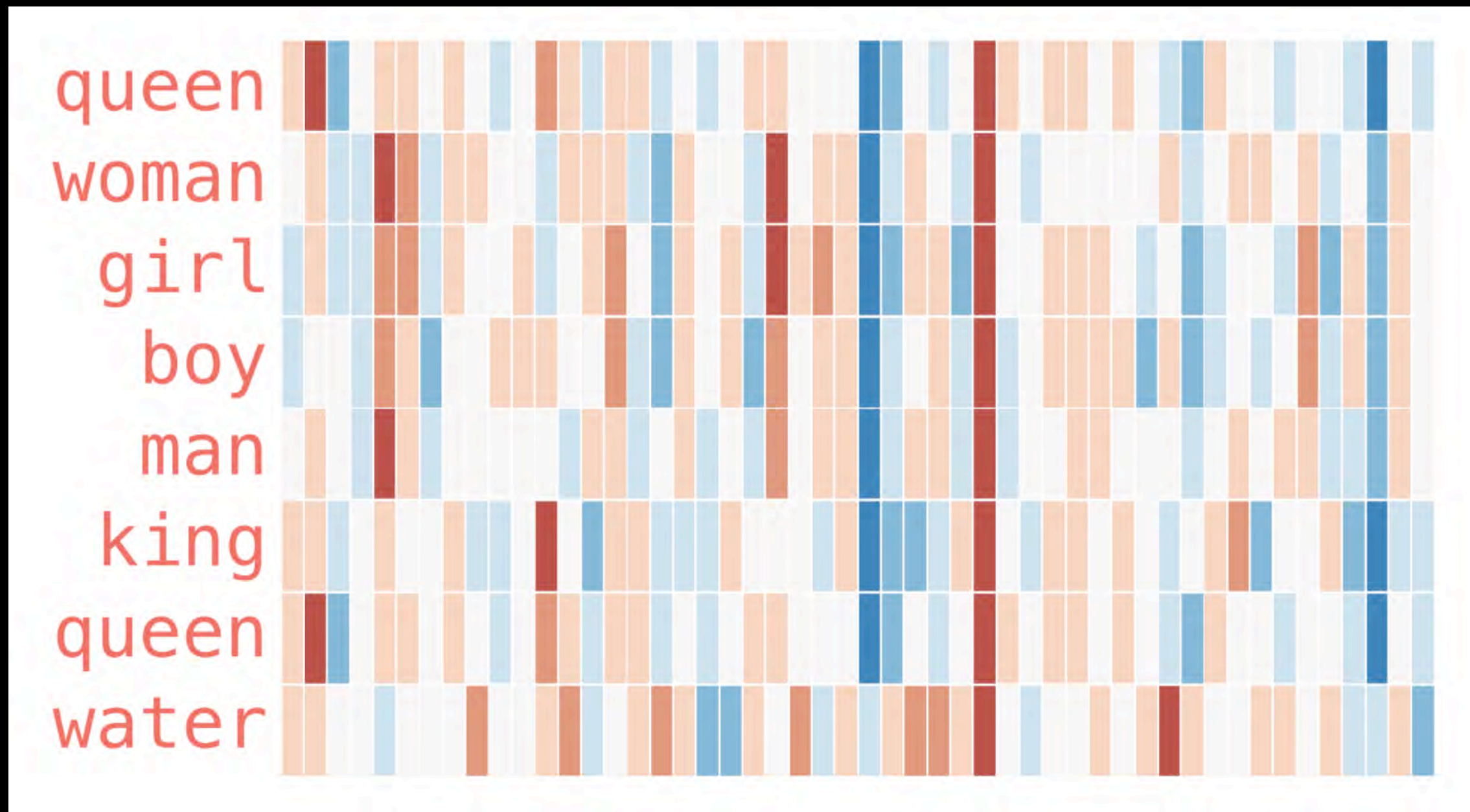


king - man + woman $\approx$ queen

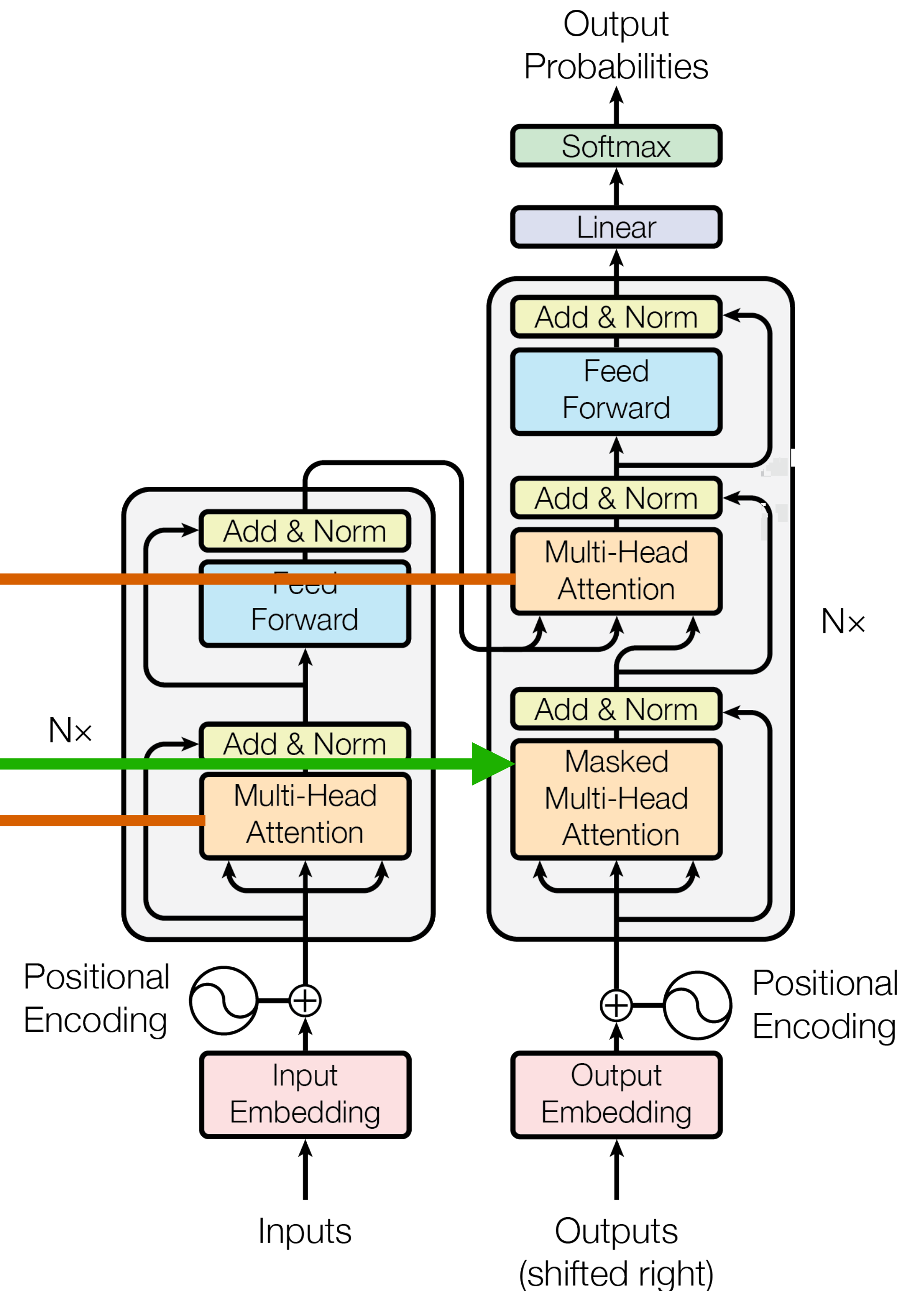


The resulting vector from "king-man+woman" doesn't exactly equal "queen", but "queen" is the closest word to it from the 400,000 word embeddings we have in this collection.

„Transformer Neural Networks“ (TNN): Grundlage von ChatGPT (z.B. BERT; GPT-3)



Alammar, J. (2019, März 27). The Illustrated Word2vec. <https://jalammar.github.io/illustrated-word2vec/>





Stable Diffusion XL, Prompt „A meeting at the rotary club“

Text zu
Bild-KI:
Logiken
und
Probleme
zukünftiger
Bildwelten

„(Latent) Diffusion Modelle“ ([L]DM)

Join our waitlist for the upcoming early preview of Stable Diffusion 3. ×

stability.ai Models ▾ Deployment ▾ Company ▾ News日本語 Contact Us

SDXL Turbo

Unleashing SDXL's power at the speed of thought.

This lightning fast model, developed by our research team and based on SDXL, is capable of generating image-based results as fast as you can type.

Try SDXL TurboDownload Code



OpenAI Menu

Blog

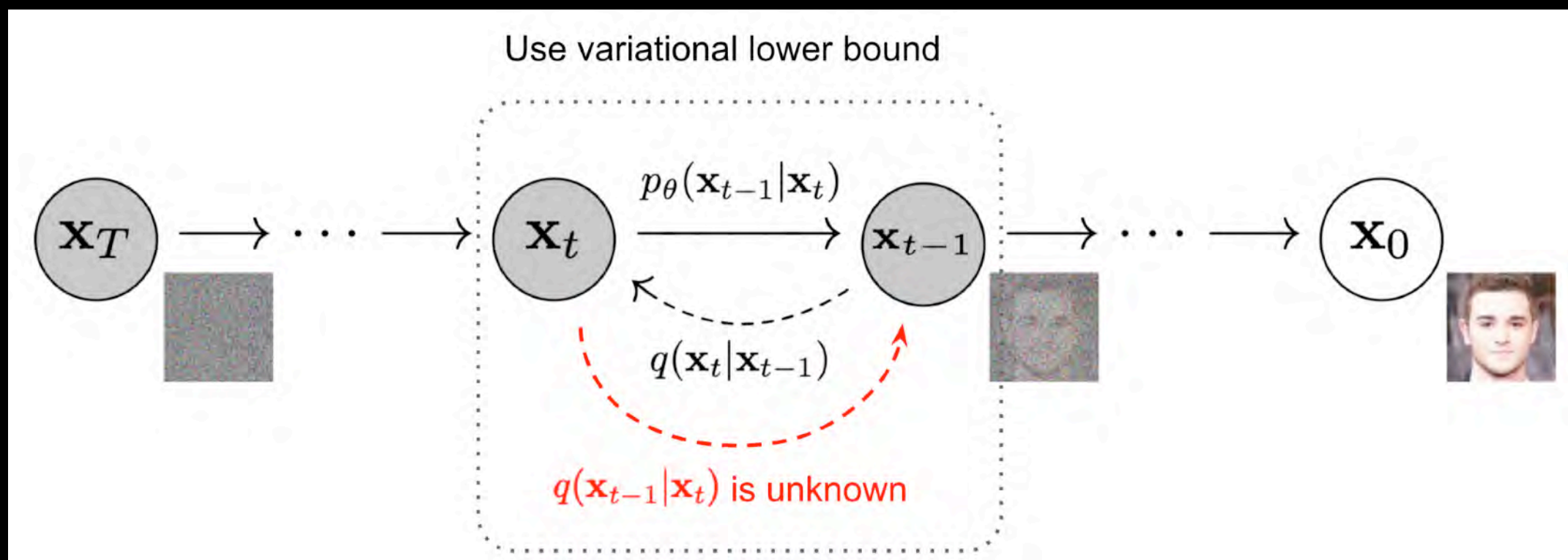
DALL·E 2: Extending creativity

As part of our DALL·E 2 research preview, more than 3,000 artists from more than 118 countries have incorporated DALL·E into their creative workflows. The artists in our early access group have helped us discover new uses for DALL·E and have served as key voices as we've made decisions about DALL·E's features.

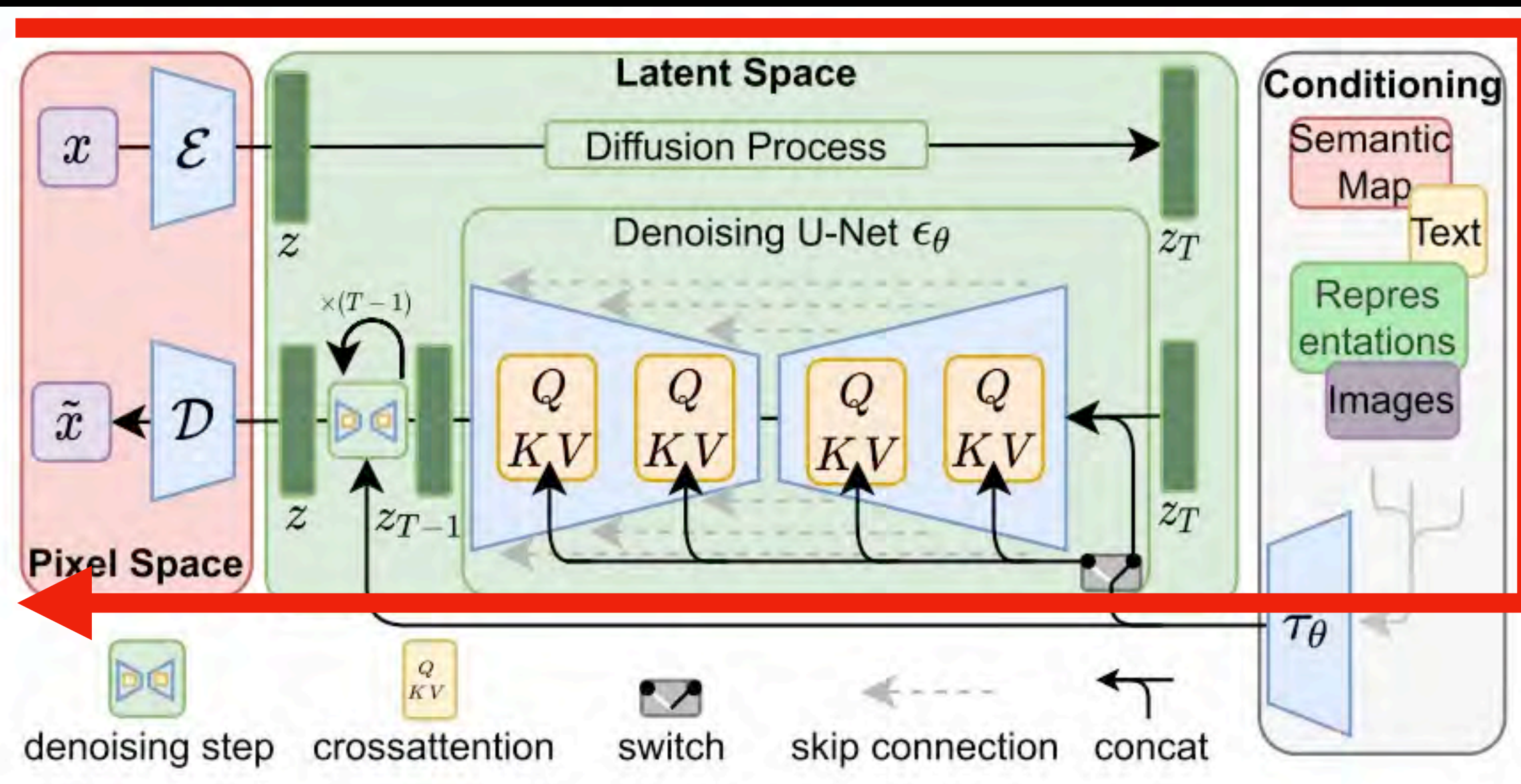


Illustration: Justin Jay Wang × DALL·E

<https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>



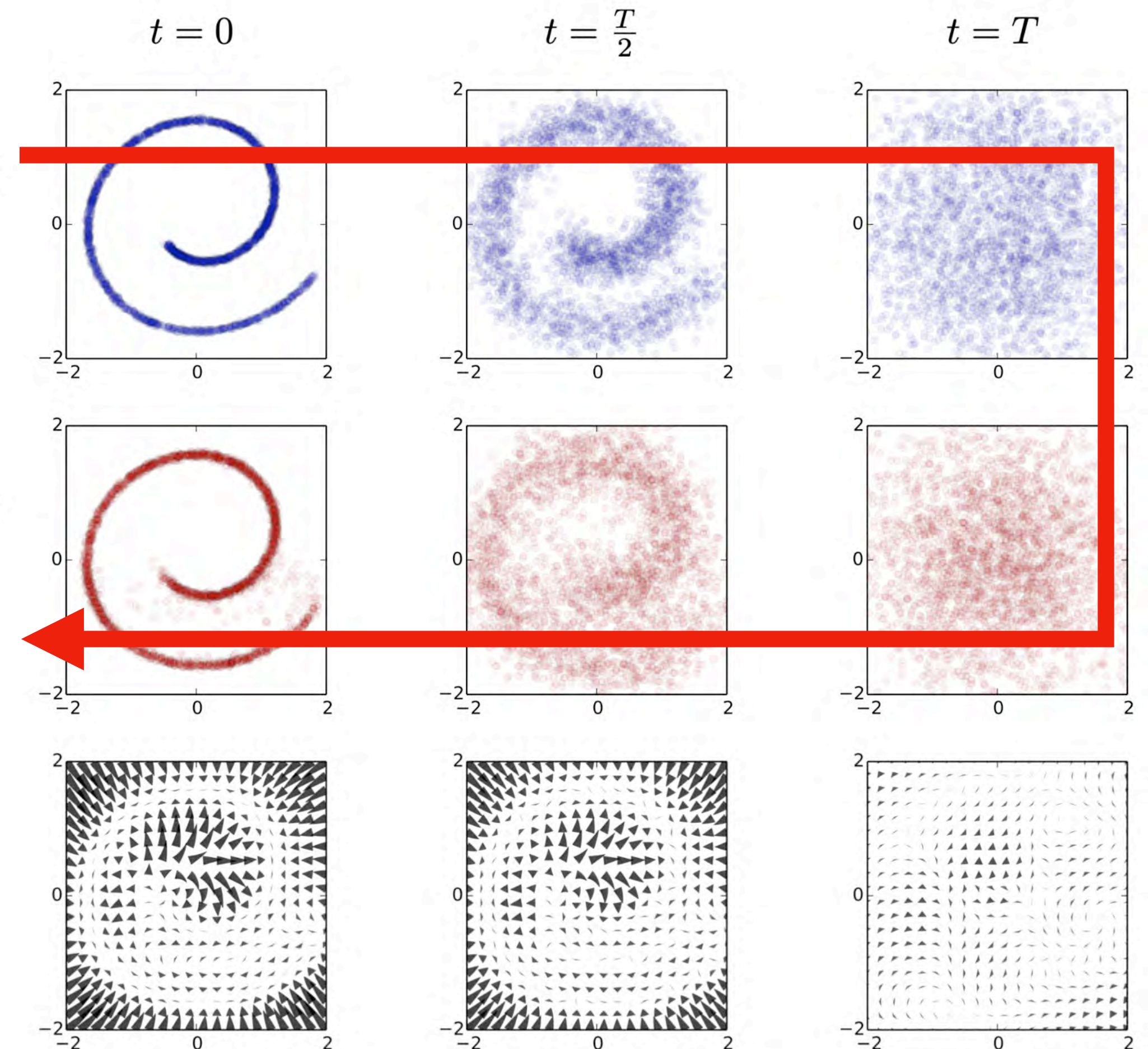
„(Latent) Diffusion Model“ ([L]DM)

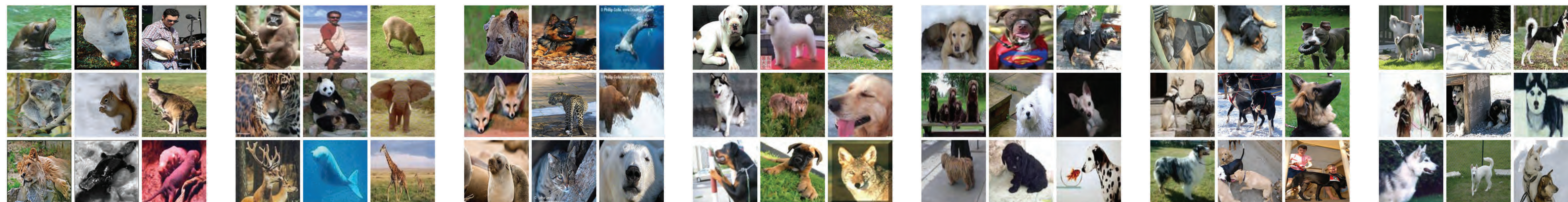


The forward trajectory
 $q(\mathbf{x}_{0:T})$

The reverse trajectory
 $p_\theta(\mathbf{x}_{0:T})$

The drifting term
 $\mu_\theta(\mathbf{x}_t, t) - \mathbf{x}_t$





mammal → placental → carnivore → canine → dog → working dog → husky



vehicle → craft → watercraft → sailing vessel

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>

WordNet Search - 3.1

- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- **S: (n) dog, domestic dog, *Canis familiaris*** (a member of the genus *Canis* (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds) *"the dog barked all night"*
- **S: (n) frump, dog** (a dull unattractive unpleasant girl or woman) *"she got a reputation as a frump"; "she's a real dog"*
- **S: (n) dog** (informal term for a man) *"you lucky dog"*
- **S: (n) cad, bounder, blackguard, dog, hound, heel** (someone who is morally reprehensible) *"you dirty dog"*
- **S: (n) frank, frankfurter, hotdog, hot dog, dog, wiener, wienerwurst, weenie** (a smooth-textured sausage of minced beef or pork usually smoked; often served on a bread roll)
- **S: (n) pawl, detent, click, dog** (a hinged catch that fits into a notch of a ratchet to move a wheel forward or prevent it from moving backward)
- **S: (n) andiron, firedog, dog, dog-iron** (metal supports for logs in a fireplace) *"the andirons were too hot to touch"*

Verb

- **S: (v) chase, chase after, trail, tail, tag, give chase, dog, go after, track** (go after with the intent to catch) *"The policeman chased the mugger down the alley"; "the dog chased the rabbit"*

Diversity of images. A third issue we identified is insufficient representation among ImageNet images. ImageNet consists of Internet images collected by querying image search engines^[1], which have been demonstrated to retrieve biased results in terms of race and gender^[4-5]. Taking gender as an example, Kay *et al.* find that when using occupations (e.g., banker) as keywords, image search results exhibit exaggerated gender ratios compared to the true real-world ratios. In addition, bias can also be introduced during the manual cleanup stage, as people are inclined to give positive responses when the given example is consistent with stereotypes^[4].

ImageNet had taken measures to diversify the images, such as keywords expansion, searching in multiple languages, and combining multiple search engines. Filtering out non-imageable synsets also mitigates the issue: with stronger visual evidence, the workers may be less prone to stereotypes. Despite these efforts, the bias in the protected attributes remains in many synsets in the person subtree. It is necessary to study how this type of bias affects models trained for downstream vision tasks, which would not be possible without high-quality annotation of image-level demographics.

To evaluate the demographics within ImageNet and propose a more representative subset of images, we annotated a set of protected attributes on images in the person subtree. We considered U.S. anti-discrimination laws which name race, color, national origin, religion, sex, gender, sexual orientation, disability, age, military history, and family status as protected attributes^[9-11]. Of these, the only potentially imageable attributes are color, gender, and age, so we proceeded to annotate these.

We follow the established protocols in annotating skin color, gender and age (details in our upcoming technical report). We annotated demographics on the 139 synsets that are considered both safe and imageable and that contain at least 100 images. We annotated 100 randomly sampled images from each synset, summing up to 13,900 images. The figure below shows the distribution of categories for different synsets, which mirrors real-world biases.

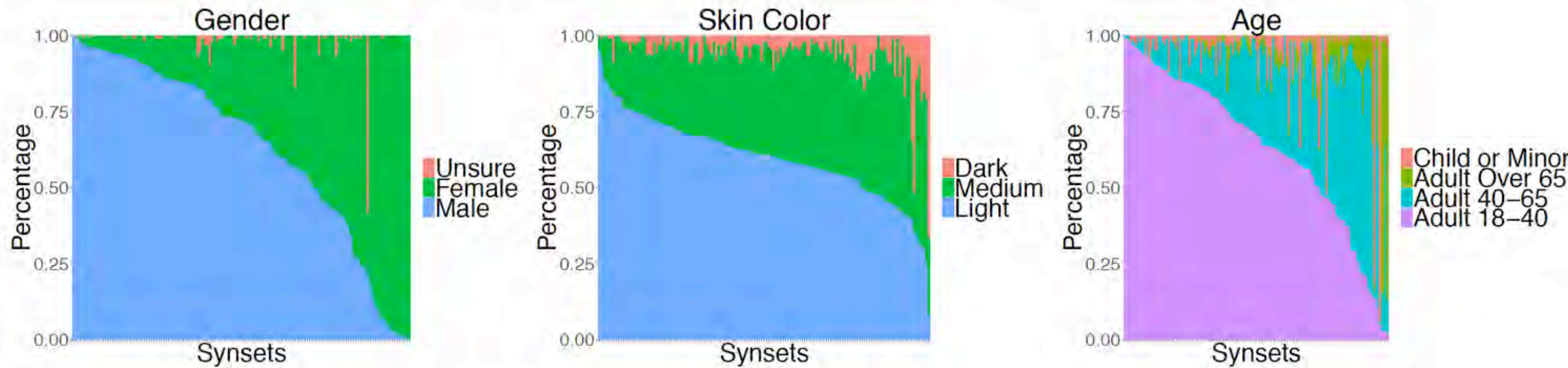


Figure 1: The distribution of demographic categories across the 139 safe and imageable synsets which contain at least 100 images. The size of the different color areas reveal the underrepresentation of certain groups.

Given the demographic analysis, it is desired to have a constructive solution to improve the diversity in ImageNet images. To this end, we are in the process of constructing a Web interface that automatically re-balances the image distribution within each synset, aiming for a target distribution of a single attribute (e.g., uniform balance of gender) by removing the images corresponding to the overrepresented categories (Figure 2). This balancing is only feasible on synsets with sufficient representation within each attribute category. For example, the synset [baby](#) naturally does not contain a balanced age distribution. Thus, we will allow the user to request a subset of the attribute categories to be balanced; for example, the user can impose equal representation of the three adult categories (“Adult Over 65”, “Adult 40-65” and “Adult 18-40”) while eliminating the “Child” category. Our future work includes collecting the rest of the annotations and publicly releasing the Web interface.

**Biases,
Sexismus,
Diskriminierung
auf Ebene der
Trainingsdatensätze**

Spanien: Jugendliche verbreiten KI-generierte Nacktbilder von Mädchen

Von fast zwei Dutzend Mädchen in Almendralejo kursierten KI-generierte Nacktbilder an ihren Schulen. 10 Jungen und eine KI-App sollen verantwortlich sein.



(Bild: BongkarnGraphic/Shutterstock.com)

22.09.2023, 14:34 Uhr Lesezeit: 3 Min.

Von Martin Holland

In Spanien sorgt der Fall von fast zwei Dutzend Mädchen für Aufsehen, von denen an ihren Schulen und unter Gleichaltrigen KI-generierte Nacktbilder verbreitet wurden. Das berichtet die Tageszeitung El País und ergänzt, dass inzwischen 22

**Biases,
Sexismus,
Diskriminierung
auf Ebene der
Trainingsdatensätze**

heise online Newsletter

Keine News verpassen! Jeden Morgen der frische

<https://www.heise.de/news/Spanien-Jugendliche-verbreiten-KI-generierte-Nacktbilder-von-Maedchen-9314057.html>

MOTHERBOARD
TECH BY VICE

ISIS Executions and Non-Consensual Porn Are Powering AI Art

AI is progressing at an astonishing speed, but we still don't have a good understanding of the datasets that power AI, and little accountability for whatever abusive images they contain.

By [Chloe Xiang](#)By [Emanuel Maiberg](#)September 21, 2022, 6:49pm [f Share](#) [t Tweet](#) [s Snap](#)

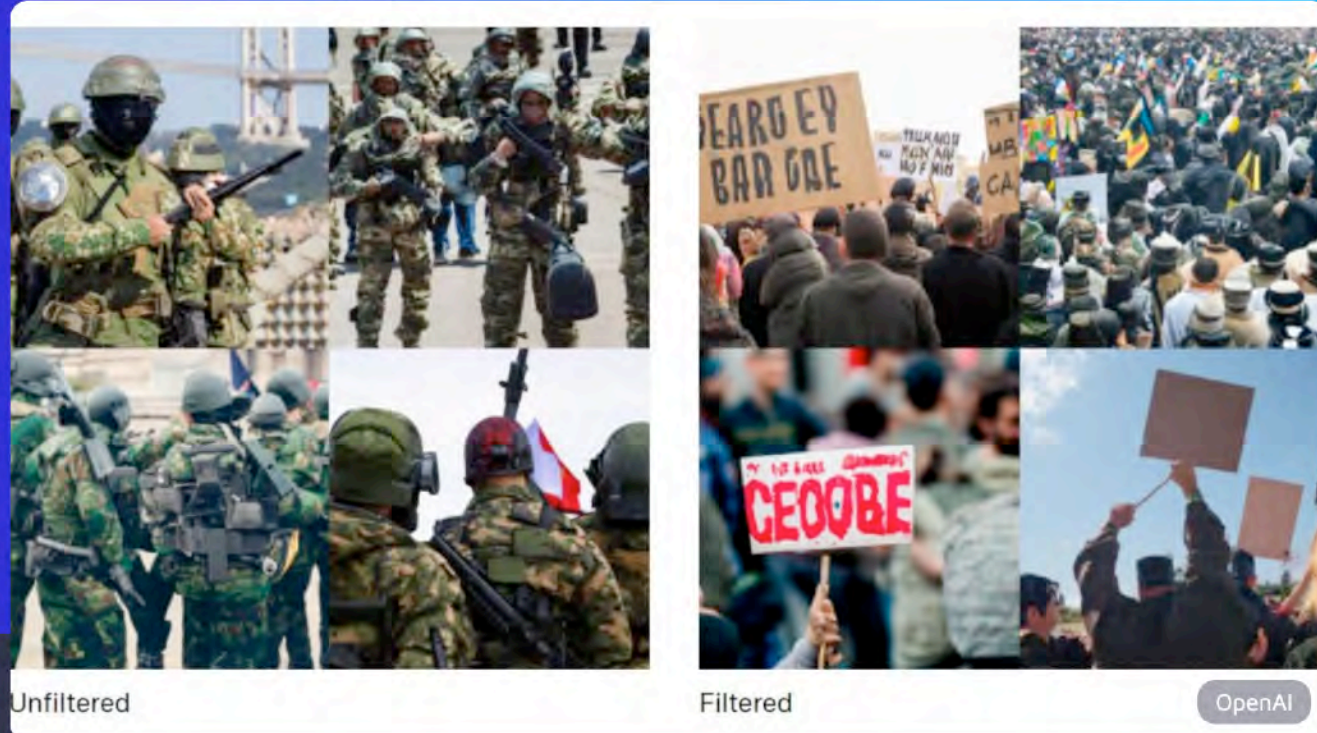
Some of the image-generating AI tools that have taken over the internet in recent months are powered in part by some of the worst images that have ever been posted to the internet, including images of the Islamic State executing people, photoshopped nudes of celebrities, and real nudes that were hacked from celebrities' phones in the 2014 incident that came to be

**Biases,
Sexismus,
Diskriminierung
auf Ebene der
Trainingsdatensätze**

[https://www.vice.com/en/article/93ad75/isis-executions-and-non-consensual-porn-are-powering-ai-art?
mc_cid=2745cdd38e&mc_eid=dcc328239e](https://www.vice.com/en/article/93ad75/isis-executions-and-non-consensual-porn-are-powering-ai-art?mc_cid=2745cdd38e&mc_eid=dcc328239e)

KI-Forschung 2. Juli 2022

OpenAI will DALL-E sicher machen - und stößt auf unerwarteten Nebeneffekt



Inhalt

Newsletter

OpenAIs DALL-E 2 setzt auf eine ganze Reihe von Sicherheitsmaßnahmen, um einen potenziellen Missbrauch zu stoppen. Jetzt gibt OpenAI einen tiefen

Datenfilter erhöht Bias im KI-Modell

Der erfolgreiche **Filterprozess** hat jedoch einen unerwarteten Nebeneffekt: Er schafft oder verstärkt den Bias des Modells gegenüber bestimmten demografischen Gruppen. Dieser Bias sei auch so eine große Herausforderung, doch der eigentlich positive Filterprozess verstärke das Problem noch einmal, so OpenAI.

Als Beispiel nennt das Unternehmen die Eingabe "ein CEO": Das ungefilterte Modell neige dazu, mehr Bilder von Männern als Frauen zu erzeugen - ein Großteil dieses Bias sei auf die Trainingsdaten zurückzuführen. Doch beim gefilterten Modell sei dieser Effekt noch verstärkt worden - es zeigte fast ausschließlich Bilder von Männern. Im Vergleich zum ungefilterten Modell sei die Frequenz des Wortes "Frau" im Datensatz um 14 Prozent reduziert, die für "Mann" lediglich sechs Prozent.

Das habe mutmaßlich zwei Ursachen: Trotz ungefähr gleicher Repräsentation von Männern und Frauen im originalen Datensatz enthalte dieser womöglich Frauen häufiger in sexualisierten Kontexten. Die Klassifizierer entfernen daher mehr Bilder von Frauen und verstärken so das Ungleichgewicht. Zusätzlich könnten die Klassifizierer selbst durch bestimmte Klassendefinitionen oder Implementierung verzerrt sein und mehr Bilder von Frauen entfernen.

OpenAI fixt Bias mit Umgewichtung der Trainingsdaten

**Biases,
Sexismus,
Diskriminierung
auf Ebene der
Trainingsdatensätze**

<https://the-decoder.de/openai-will-dall-e-sicher-machen-und-stoesst-auf-unerwarteten-nebeneffekt/>



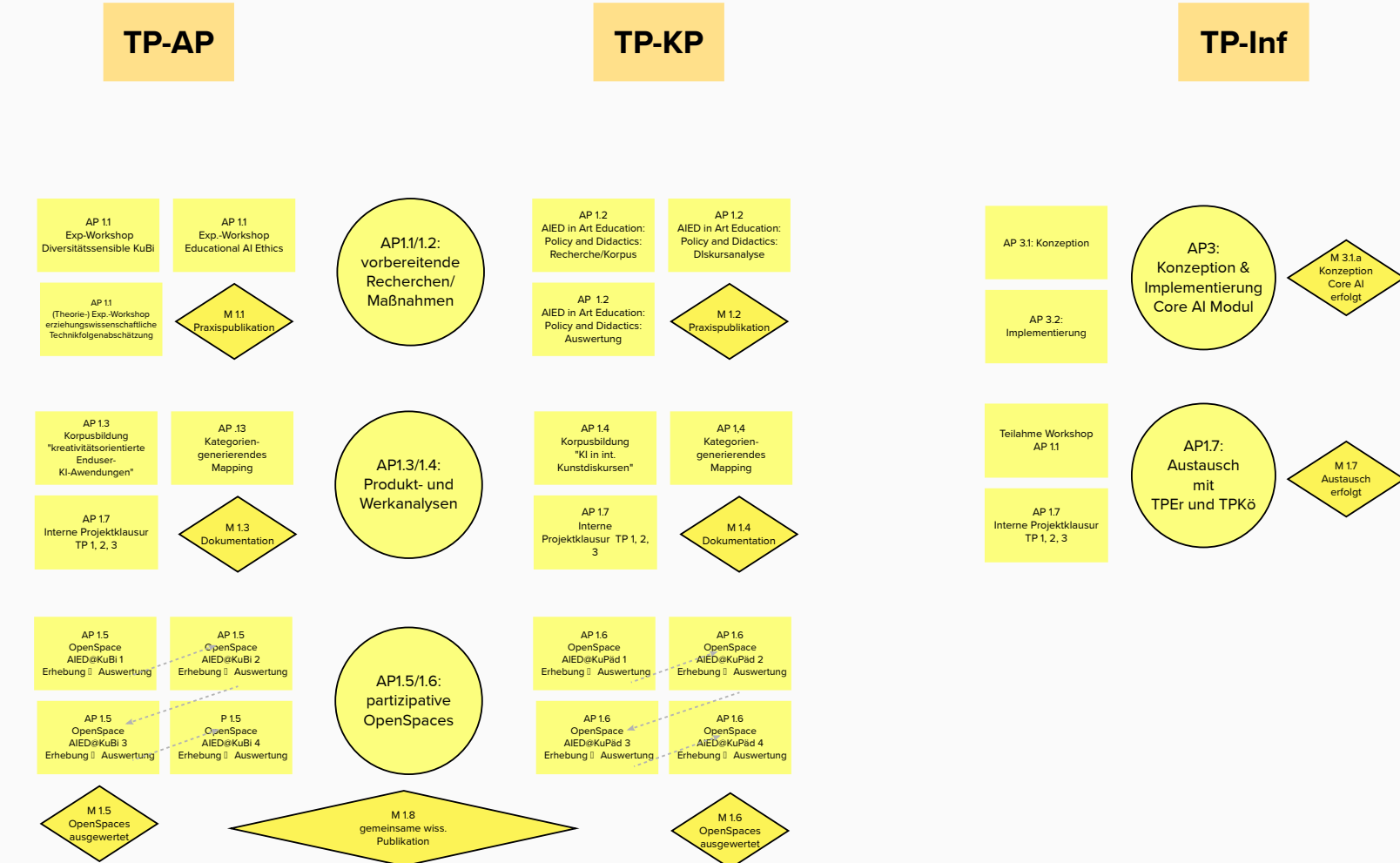
ARTIFICIAL INTELLIGENCE FOR ARTS EDUCATION (AI4ARTSED)



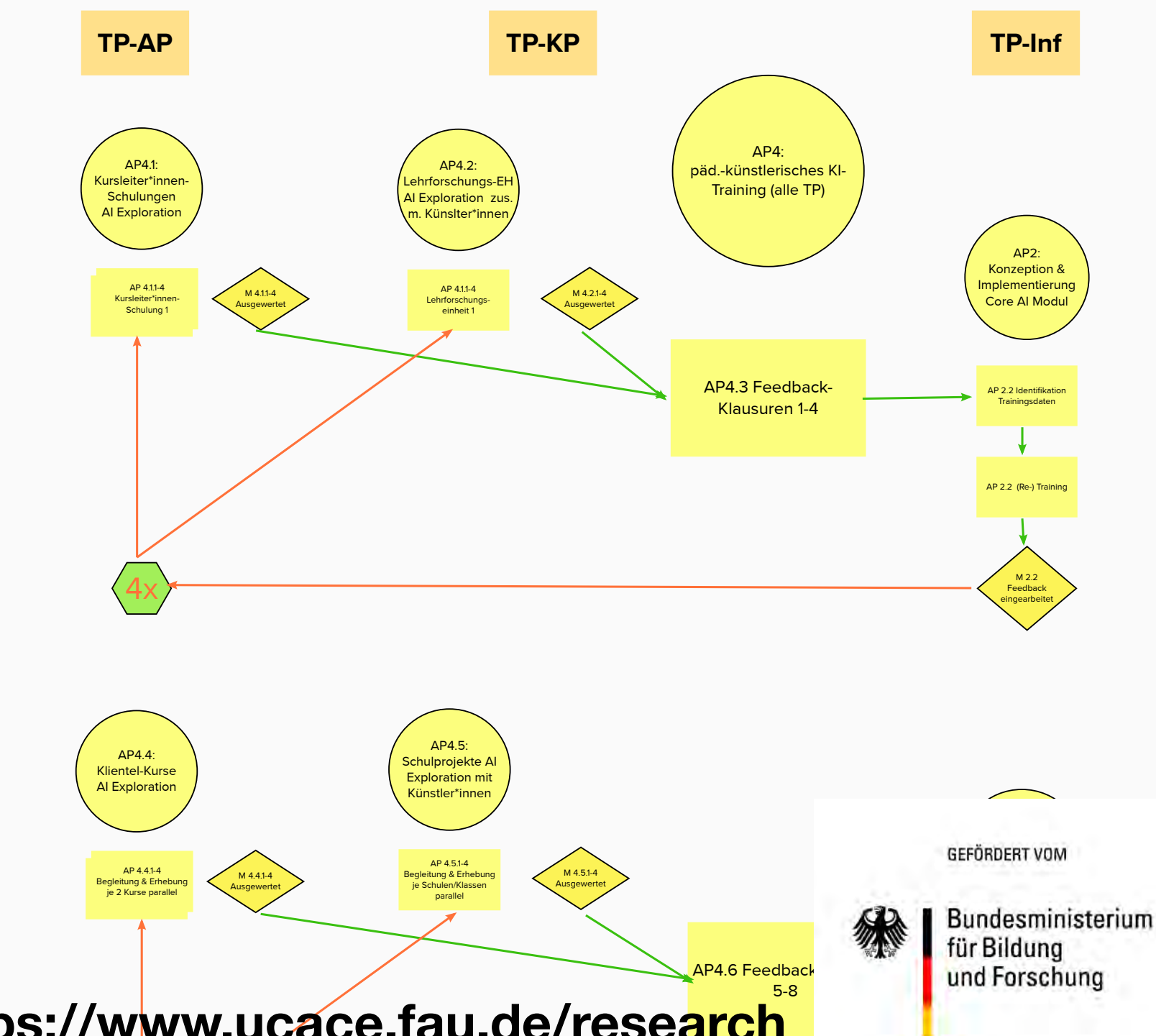
ARTIFICIAL INTELLIGENCE FOR ARTS EDUCATION – AI4ARTSED

KI verändert Gesellschaft und Arbeitswelt; sie wird zunehmend Thema der Bildung. Das Projekt sondiert die Chancen, Einsatzbedingungen und strukturellen Grenzen des Einsatzes partizipatorisch ausgerichteter künstlicher Intelligenz (KI) in kulturell diversitätssensiblen künstlerisch-pädagogischen Settings der Kulturellen Bildung. Welche und Chancen, Bedingungen und Grenzen des pädagogischen Einsatzes künstlicher Intelligenz (KI) bestehen in kulturell diversitätssensiblen Settings der Kulturellen Bildung (KuBi)? In drei Teilprojekten – Allgemeinpädagogik (TPap), Informatik (TPinf) und Kunstpädagogik (TPkp) – greifen kreativitätsorientierte pädagogische KI-Praxisforschung und informatische KI-Konzeption und Pro-

2024



2025



<https://www.ucace.fau.de/research>

GEFÖRDERT VOM



Bundesministerium für Bildung und Forschung

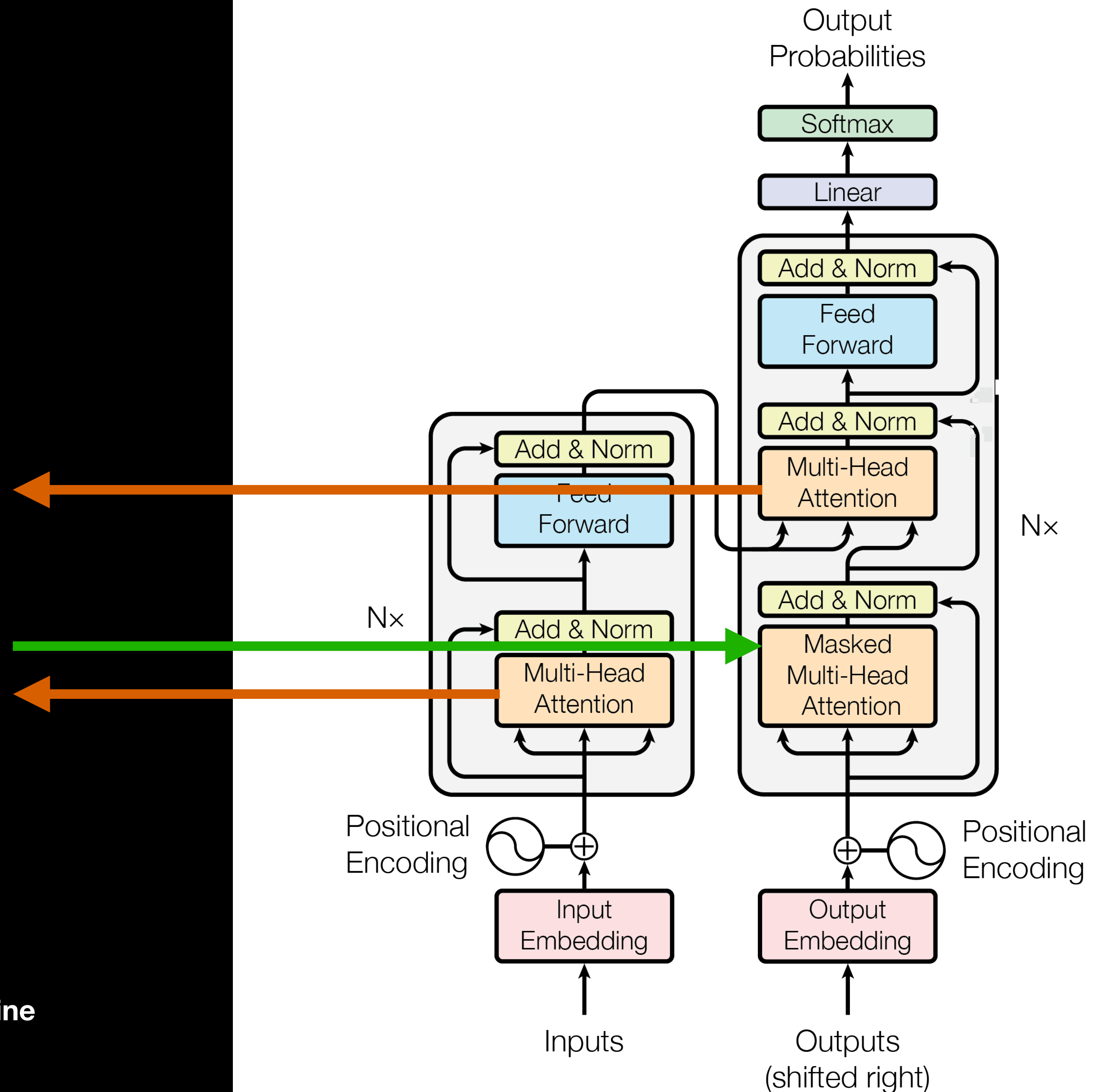


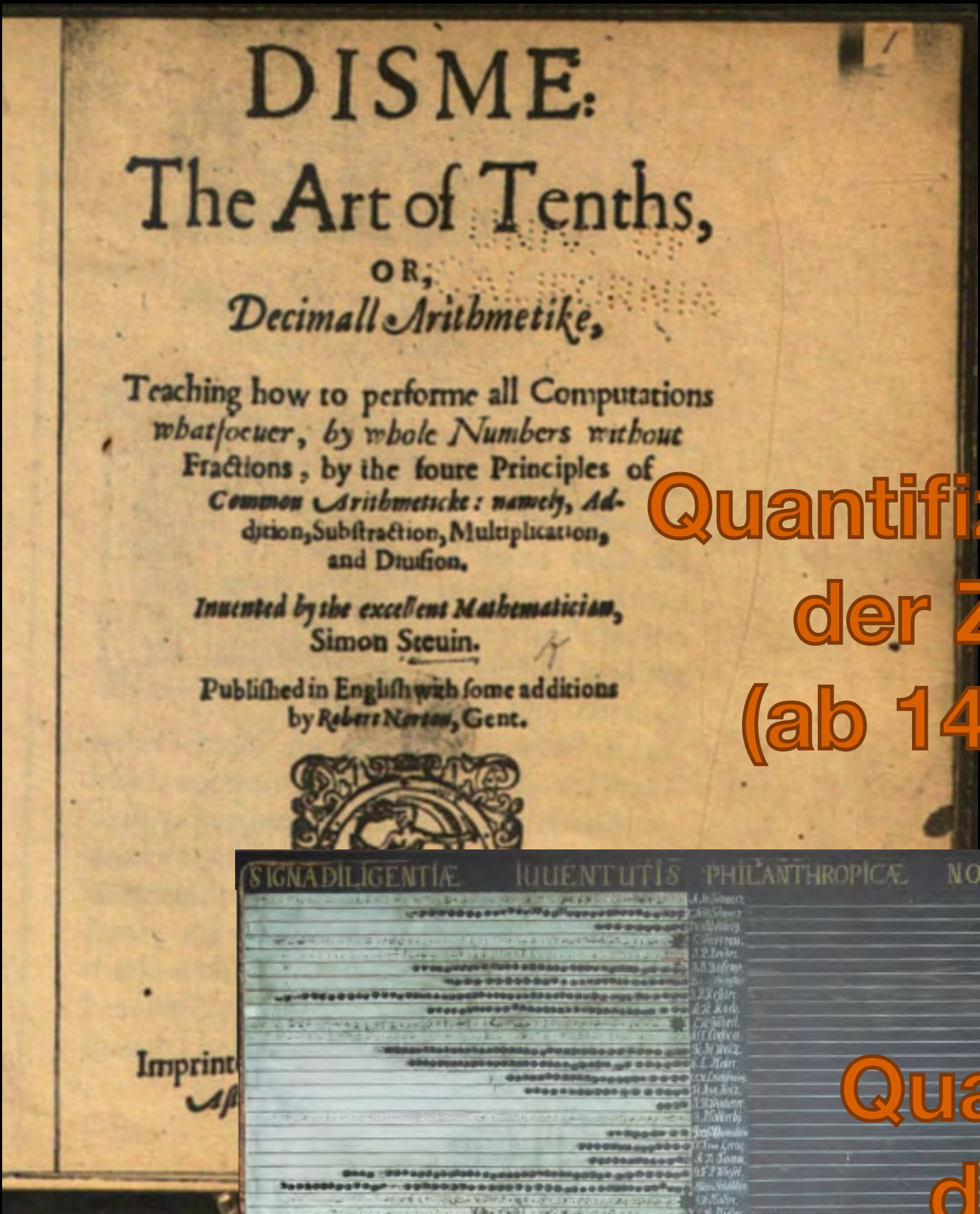
Algorithmische
Rationalität
oder „alien
space of
reasoning“?

Generelles Merkmal von
KI-Technologien im Kontext
humaner/hermeneutischer
Bedeutungswelten:

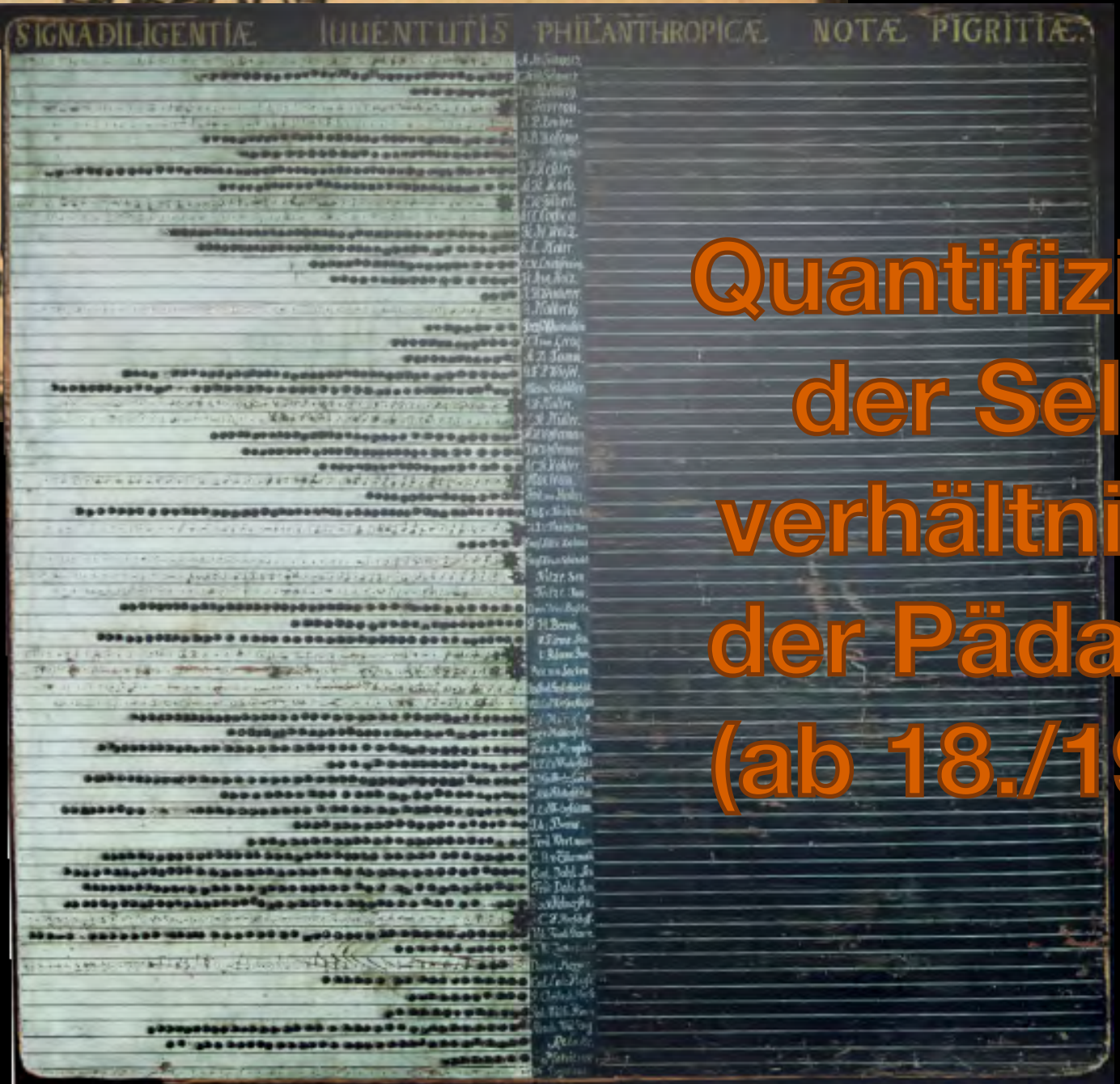
Big Data-basierte statistische
Normierung von Bedeutung:
quantitative „Wahrscheinlichkeit“
vs. verhältnismäßige
„Wahr-Scheinlichkeit“

Böhme, G. (1996). Idee und Kosmos: Platons Zeitlehre: eine
Einführung in seine theoretische Philosophie.
Klostermann, S. 18

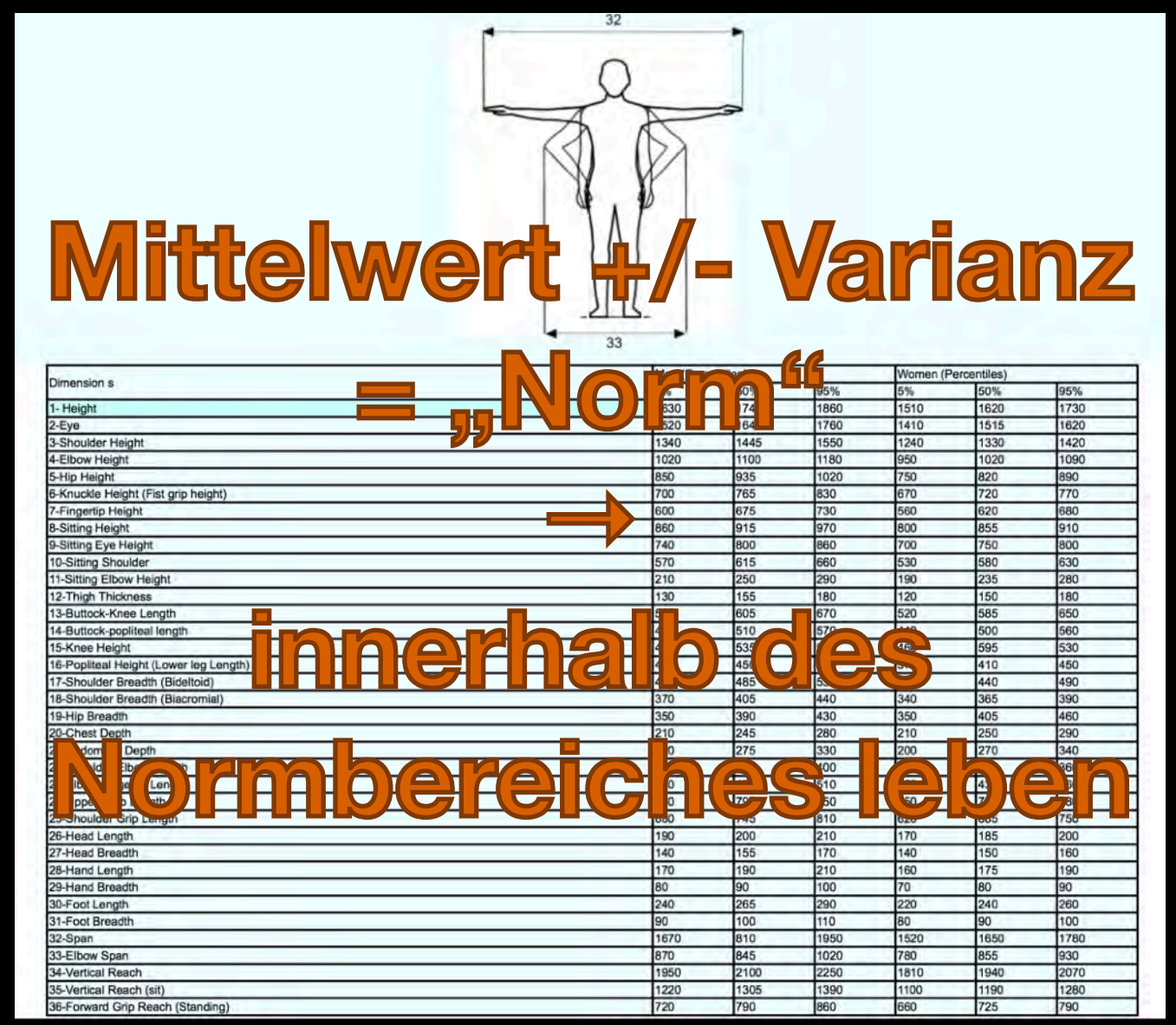




Quantifizierung
der Zahl
(ab 14. Jh.)



Quantifizierung
der Selbst-
verhältnisse &
der Pädagogik
(ab 18./19. Jh.)



Mittelwert +/- Varianz
= „Norm“
→ innerhalb des
Normbereiches leben

„Wahrscheinlichkeit“
„probability“



Proportion =
„ratio“
→ „verhältnismäßig
leben“

„likeness“
„eikos logos“
„Wahr-Scheinlichkeit“

Generelles Merkmal von
KI-Technologien im Kontext
humaner/hermeneutischer
Bedeutungswelten:

Big Data-basierte statistische
Normierung von Bedeutung:
quantitative „Wahrscheinlichkeit“
vs. verhältnismäßige
„Wahr-Scheinlichkeit“

Das Problem liegt *nicht* in den
Algorithmen/der KI an sich, sondern ...

- 1) *Produktion*: Versuch, **die Simulation von Bedeutung durch Statistik als User Experience zu installieren**, d.h. zu maskieren: User Experience ↔ Compliance als Geschäftsmodell
- 2) *Rezeption*: **Anthropomorphisierung der Maschine** (Verwechslung von Wahrscheinlichkeit und „Wahr-Scheinlichkeit“)
- 3) *Politik + Wirtschaft*: **Universalisierung und Automatisierung kybernetischer Steuerungslogiken** (Social Engineering, Subjektengineering, Humanengineering, Militär, Polizei, Rechtssprechung)

Generelles Merkmal von
KI-Technologien im Kontext
humaner/hermeneutischer
Bedeutungswelten:

Big Data-basierte statistische
Normierung von Bedeutung:

quantitative „Wahrscheinlichkeit“
vs. verhältnismäßige
„Wahr-Scheinlichkeit“

→ zu unterscheidende Ebenen von
Analyse und Kritik:

- 1) **Kritische Analytik/Hermeneutik von KI** auf der Ebene ihrer operativen Logiken (folgender Abschnitt)
 - 1) KI als „andere“, non-humane Intelligenz, aber welche Art von Alterität mit welchen Implikationen?
- 2) **Kritische Rekonstruktion der Praktiken und Policies** in Bezug auf Hervorbringung von und Umgang mit KI, d.h. v.a. ihrer politischen Ökonomie
 - 1) Exklusionen (Datenformate, Daten, Archive)
 - 2) Politiken und Normativitäten, implizite Ideologien (Ground truths, Firmenpolicies des Trainings und Finetunings)

Mathematik (Mersch)	KI (Mersch)
nicht durchgängig computierbar	universelle Computierbarkeit

Mersch, D. (2019). Kreativität und Künstliche Intelligenz. Einige Bemerkungen zu einer Kritik algorithmischer Rationalität. Zeitschrift für Medienwissenschaft, 11(2), 65–74. <https://doi.org/10.25969/mediarep/12634>

Mathematik (Mersch)	KI (Mersch)
nicht durchgängig computierbar	universelle Computierbarkeit
Poetik der Findungen	vollständige Berechenbarkeit

Mersch, D. (2019). Kreativität und Künstliche Intelligenz. Einige Bemerkungen zu einer Kritik algorithmischer Rationalität. Zeitschrift für Medienwissenschaft, 11(2), 65–74. <https://doi.org/10.25969/mediarep/12634>

Mathematik (Mersch)	KI (Mersch)
nicht durchgängig computierbar	universelle Computierbarkeit
Poetik der Findungen	vollständige Berechenbarkeit
Differenz von Berechenbarkeit und Nichtberechenbarkeit	deduktive formale Geschlossenheit

Mersch, D. (2019). Kreativität und Künstliche Intelligenz. Einige Bemerkungen zu einer Kritik algorithmischer Rationalität. Zeitschrift für Medienwissenschaft, 11(2), 65–74. <https://doi.org/10.25969/mediarep/12634>

Mathematik (Mersch)	KI (Mersch)
nicht durchgängig computierbar	universelle Computierbarkeit
Poetik der Findungen	vollständige Berechenbarkeit
Differenz von Berechenbarkeit und Nichtberechenbarkeit	deduktive formale Geschlossenheit
strukturell unvollständig	algorithmische Rationalität

Mersch, D. (2019). Kreativität und Künstliche Intelligenz. Einige Bemerkungen zu einer Kritik algorithmischer Rationalität. Zeitschrift für Medienwissenschaft, 11(2), 65–74. <https://doi.org/10.25969/mediarep/12634>

Mathematik (Mersch)	KI (Mersch)	KI (Parisi)
nicht durchgängig computierbar	universelle Computierbarkeit	Unverfügbarkeit als grundsätzliches, immanentes Moment digitaler (Turing-) Maschinen
Poetik der Findungen	vollständige Berechenbarkeit	
Differenz von Berechenbarkeit und Nichtberechenbarkeit	deduktive formale Geschlossenheit	
strukturell unvollständig	algorithmische Rationalität	

Mersch, D. (2019). Kreativität und Künstliche Intelligenz. Einige Bemerkungen zu einer Kritik algorithmischer Rationalität. Zeitschrift für Medienwissenschaft, 11(2), 65–74. <https://doi.org/10.25969/mediarep/12634>

Parisi, L. (2019). The alien subject of AI. Subjectivity, 12(1), 27–48. <https://doi.org/10.1057/s41286-018-00064-3>

Mathematik (Mersch)	KI (Mersch)	KI (Parisi)
nicht durchgängig computierbar	universelle Computierbarkeit	Unverfügbarkeit als grundsätzliches, immanentes Moment digitaler (Turing-) Maschinen
Poetik der Findungen	vollständige Berechenbarkeit	„Halteproblem“; „Zufälle/Unfälle und Fehler“ als integraler „Bestandteil der interaktiven laufenden Arbeit mit kollidierenden Daten“
Differenz von Berechenbarkeit und Nichtberechenbarkeit	deduktive formale Geschlossenheit	
strukturell unvollständig	algorithmische Rationalität	

Mersch, D. (2019). Kreativität und Künstliche Intelligenz. Einige Bemerkungen zu einer Kritik algorithmischer Rationalität. Zeitschrift für Medienwissenschaft, 11(2), 65–74. <https://doi.org/10.25969/mediarep/12634>

Parisi, L. (2019). The alien subject of AI. Subjectivity, 12(1), 27–48. <https://doi.org/10.1057/s41286-018-00064-3>

Mathematik (Mersch)	KI (Mersch)	KI (Parisi)
nicht durchgängig computierbar	universelle Computierbarkeit	Unverfügbarkeit als grundsätzliches, immanentes Moment digitaler (Turing-) Maschinen
Poetik der Findungen	vollständige Berechenbarkeit	„Halteproblem“; „Zufälle/Unfälle und Fehler“ als integraler „Bestandteil der interaktiven laufenden Arbeit mit kollidierenden Daten“
Differenz von Berechenbarkeit und Nichtberechenbarkeit	deduktive formale Geschlossenheit	affektgeladen durch menschengemachte Daten
strukturell unvollständig	algorithmische Rationalität	

Mersch, D. (2019). Kreativität und Künstliche Intelligenz. Einige Bemerkungen zu einer Kritik algorithmischer Rationalität. Zeitschrift für Medienwissenschaft, 11(2), 65–74. <https://doi.org/10.25969/mediarep/12634>

Parisi, L. (2019). The alien subject of AI. Subjectivity, 12(1), 27–48. <https://doi.org/10.1057/s41286-018-00064-3>

Mathematik (Mersch)	KI (Mersch)	KI (Parisi)
nicht durchgängig computierbar	universelle Computierbarkeit	Unverfügbarkeit als grundsätzliches, immanentes Moment digitaler (Turing-) Maschinen
Poetik der Findungen	vollständige Berechenbarkeit	„Halteproblem“; „Zufälle/Unfälle und Fehler“ als integraler „Bestandteil der interaktiven laufenden Arbeit mit kollidierenden Daten“
Differenz von Berechenbarkeit und Nichtberechenbarkeit	deduktive formale Geschlossenheit	affektgeladen durch menschengemachte Daten
strukturell unvollständig	algorithmische Rationalität	„alien space of reasoning“; Hegemonialität nicht grundsätzlich, sondern Folge hegenomnialer Praxis

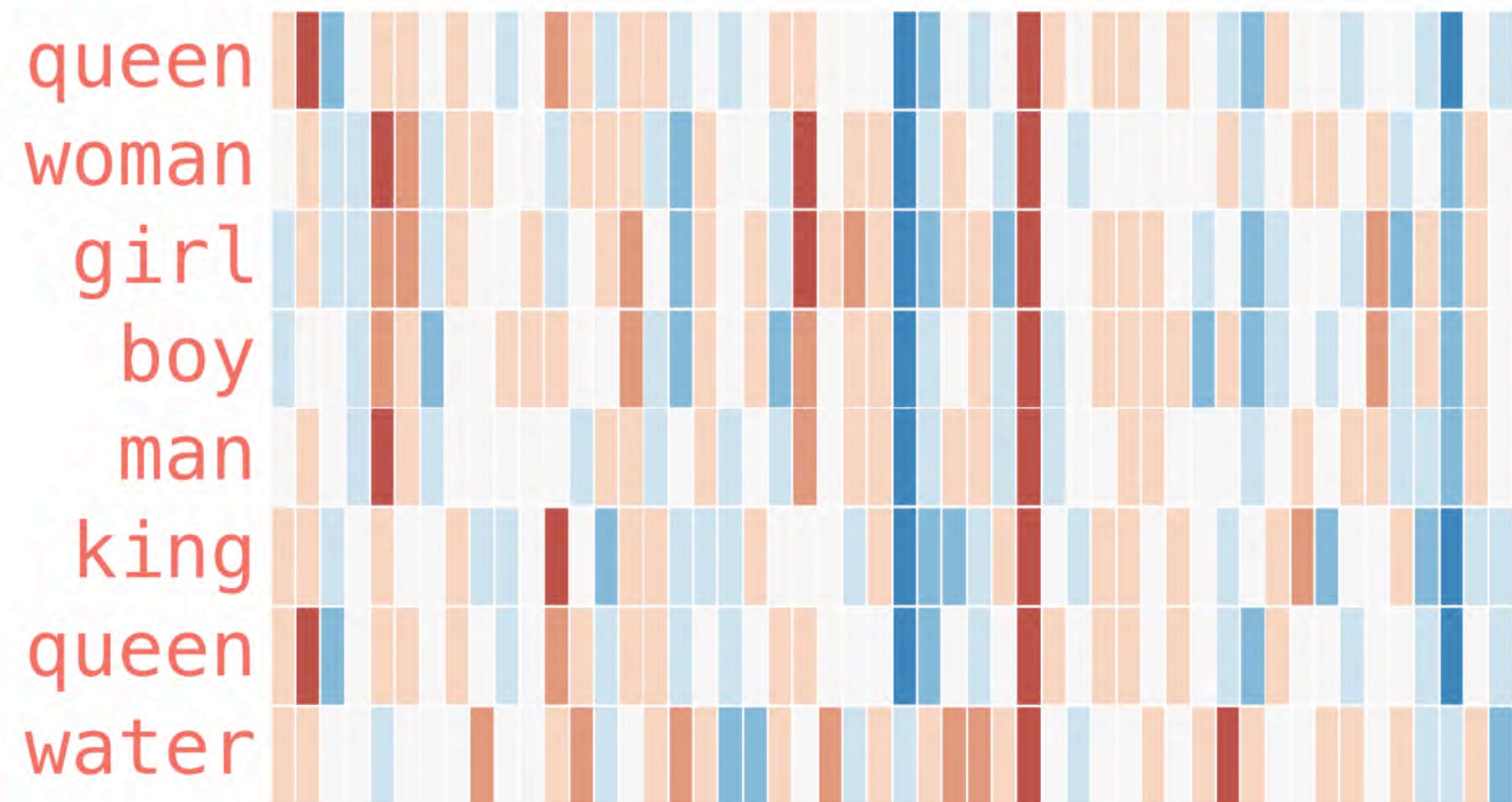
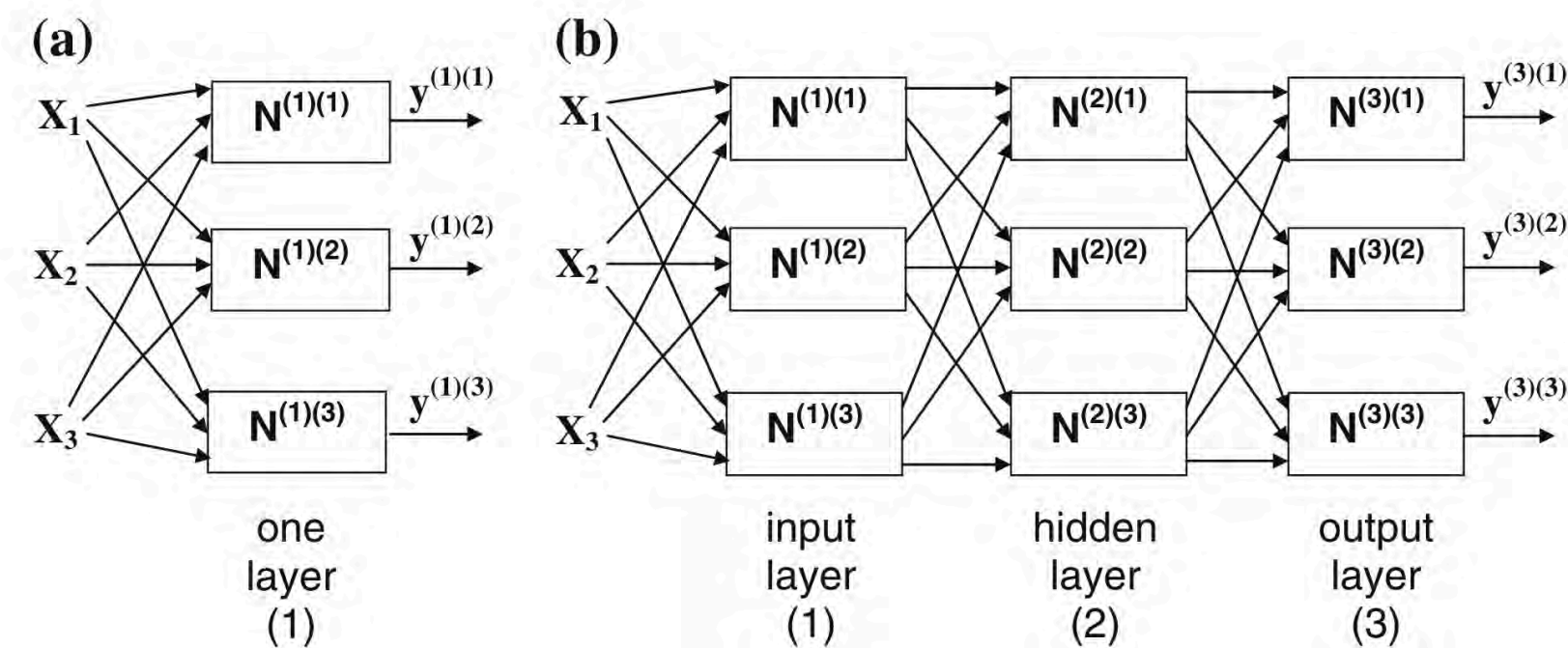
Mersch, D. (2019). Kreativität und Künstliche Intelligenz. Einige Bemerkungen zu einer Kritik algorithmischer Rationalität. Zeitschrift für Medienwissenschaft, 11(2), 65–74. <https://doi.org/10.25969/mediarep/12634>

Parisi, L. (2019). The alien subject of AI. Subjectivity, 12(1), 27–48. <https://doi.org/10.1057/s41286-018-00064-3>

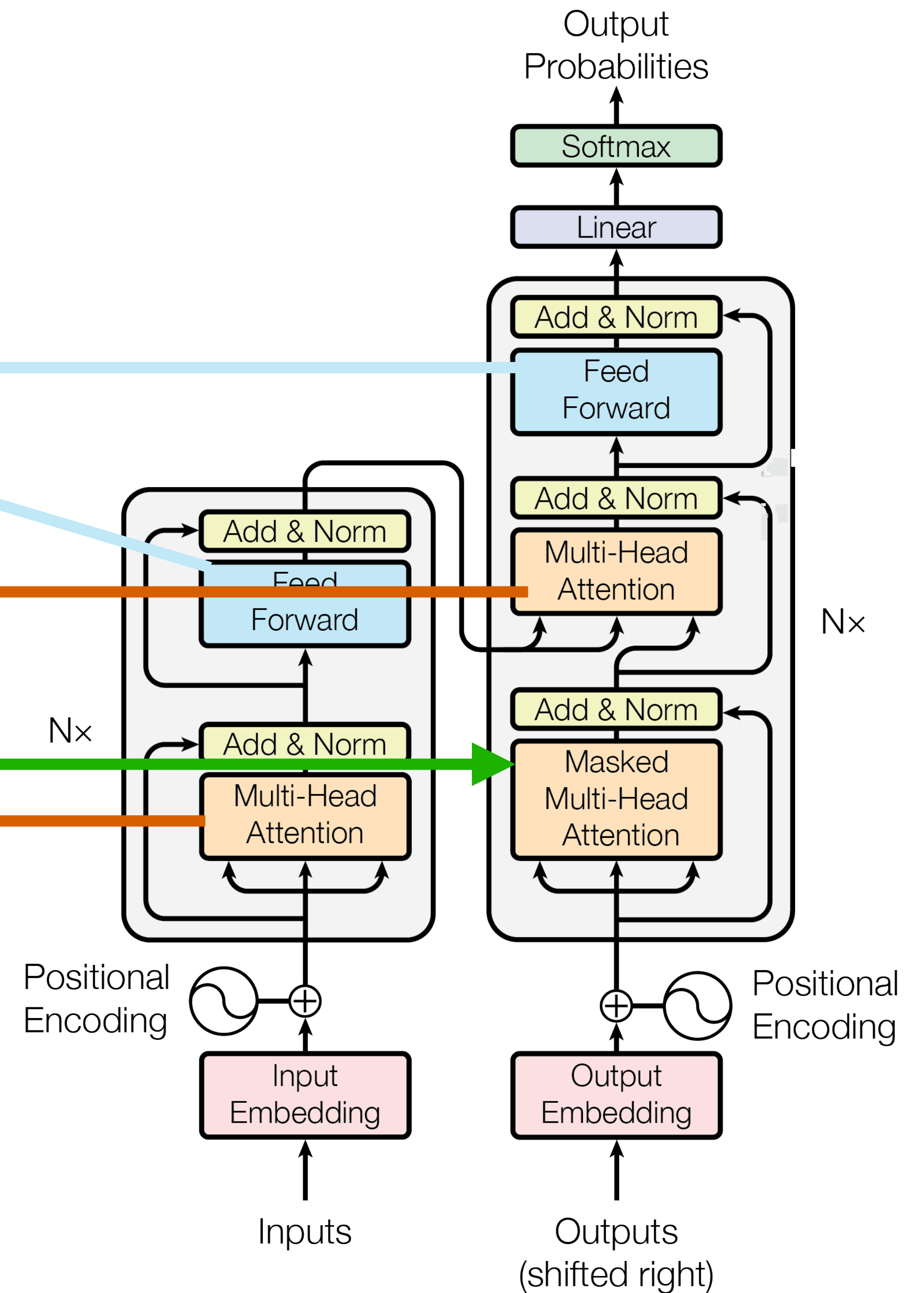
Mathematik (Mersch)	KI (Mersch)	KI (Parisi)
nicht durchgängig computierbar	universelle Computierbarkeit	Unverfügbarkeit als grundsätzliches, immanentes Moment digitaler (Turing-) Maschinen
Poetik der Findungen	vollständige Berechenbarkeit	„Halteproblem“; „Zufälle/Unfälle und Fehler“ als integraler „Bestandteil der interaktiven laufenden Arbeit mit kollidierenden Daten“
Differenz von Berechenbarkeit und Nichtberechenbarkeit	deduktive formale Geschlossenheit	affektgeladen durch menschengemachte Daten
strukturell unvollständig	algorithmische Rationalität	„alien space of reasoning“; Hegemonialität nicht grundsätzlich, sondern Folge hegenomnialer Praxis
schöpferisches Tätigkeitsfeld	unkreatives, totales Kontrollparadigma	kreativer Kontrollverlust (wo nicht hegemonial eingehegt)

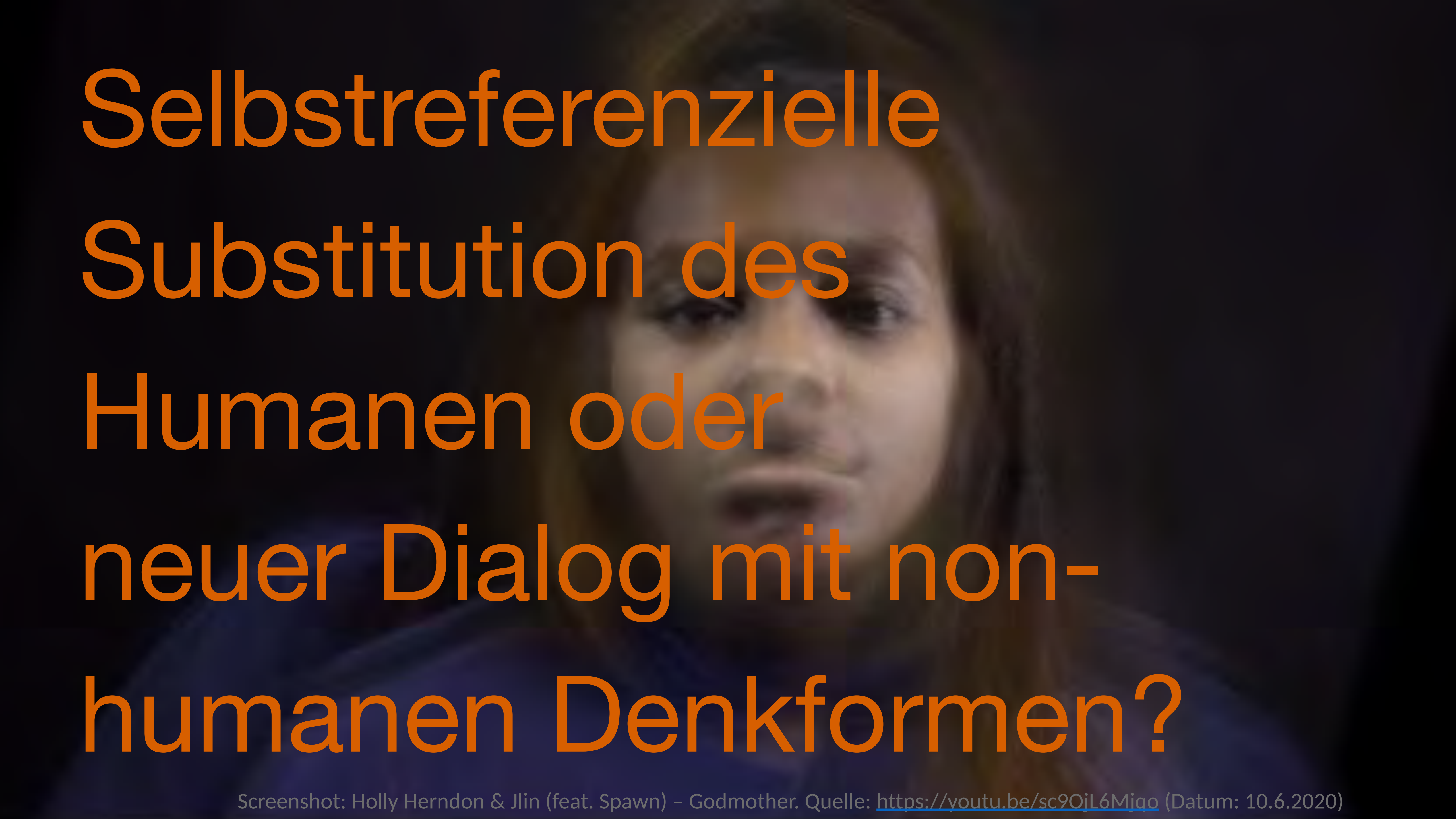
Mersch, D. (2019). Kreativität und Künstliche Intelligenz. Einige Bemerkungen zu einer Kritik algorithmischer Rationalität. Zeitschrift für Medienwissenschaft, 11(2), 65–74. <https://doi.org/10.25969/mediarep/12634>

Parisi, L. (2019). The alien subject of AI. Subjectivity, 12(1), 27–48. <https://doi.org/10.1057/s41286-018-00064-3>

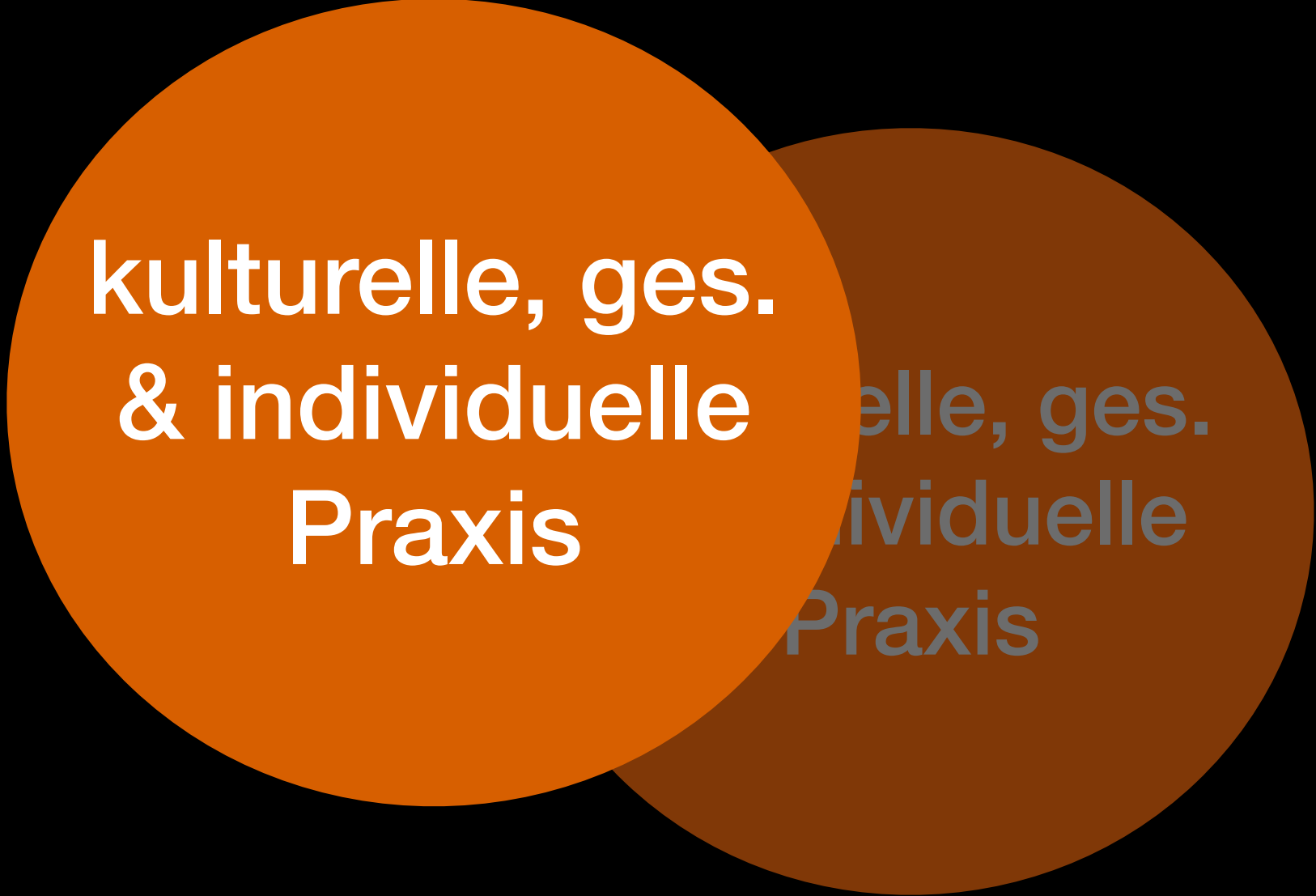


Alammar, J. (2019, März 27). The Illustrated Word2vec. <https://jalammar.github.io/illustrated-word2vec/>



A blurred background image of a woman's face, likely Holly Herndon, with her eyes looking slightly to the side. The image is dark and out of focus, serving as a backdrop for the text.

Selbstreferenzielle
Substitution des
Humanen oder
neuer Dialog mit non-
humanen Denkformen?



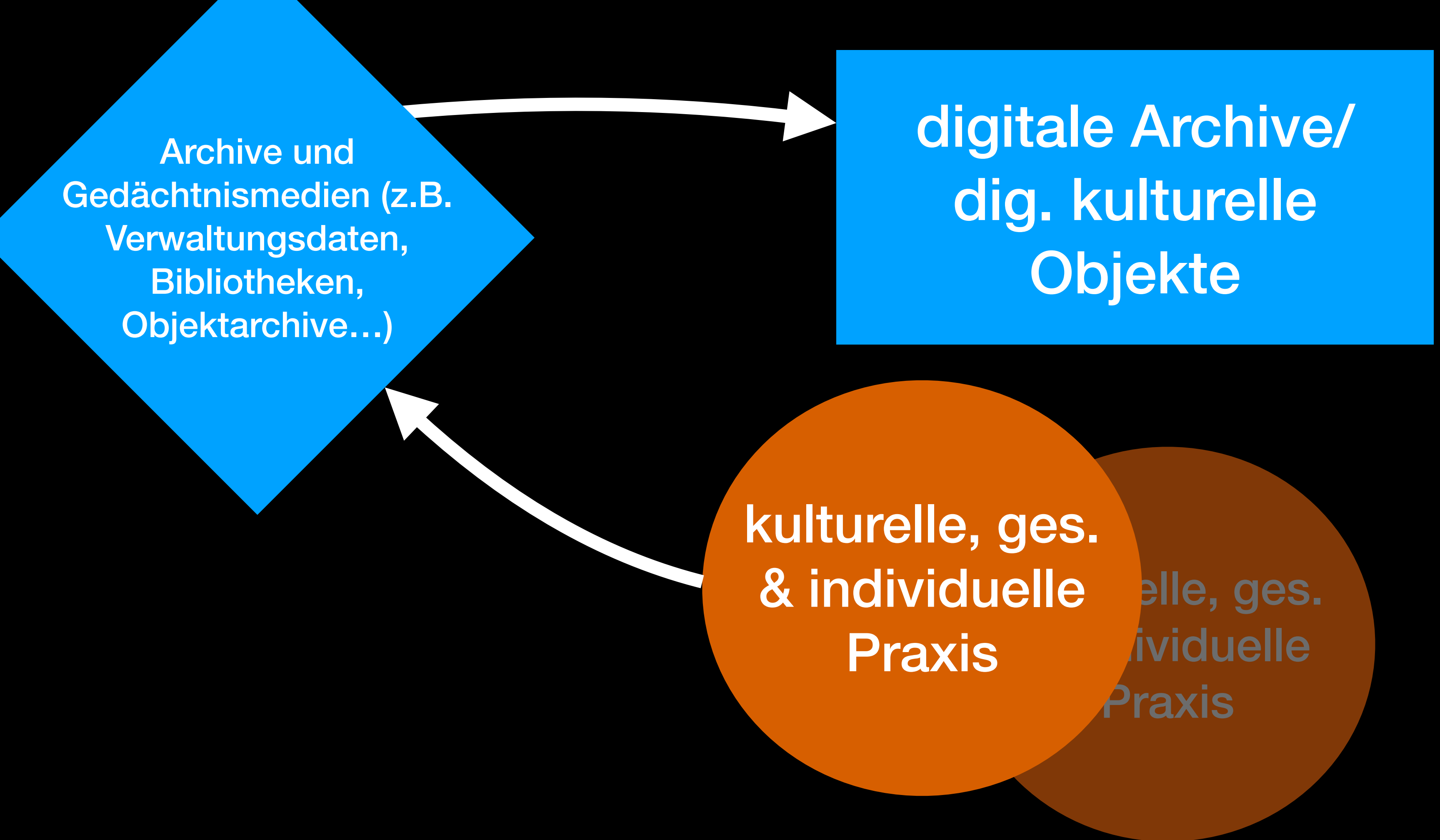
kulturelle, ges.
& individuelle
Praxis

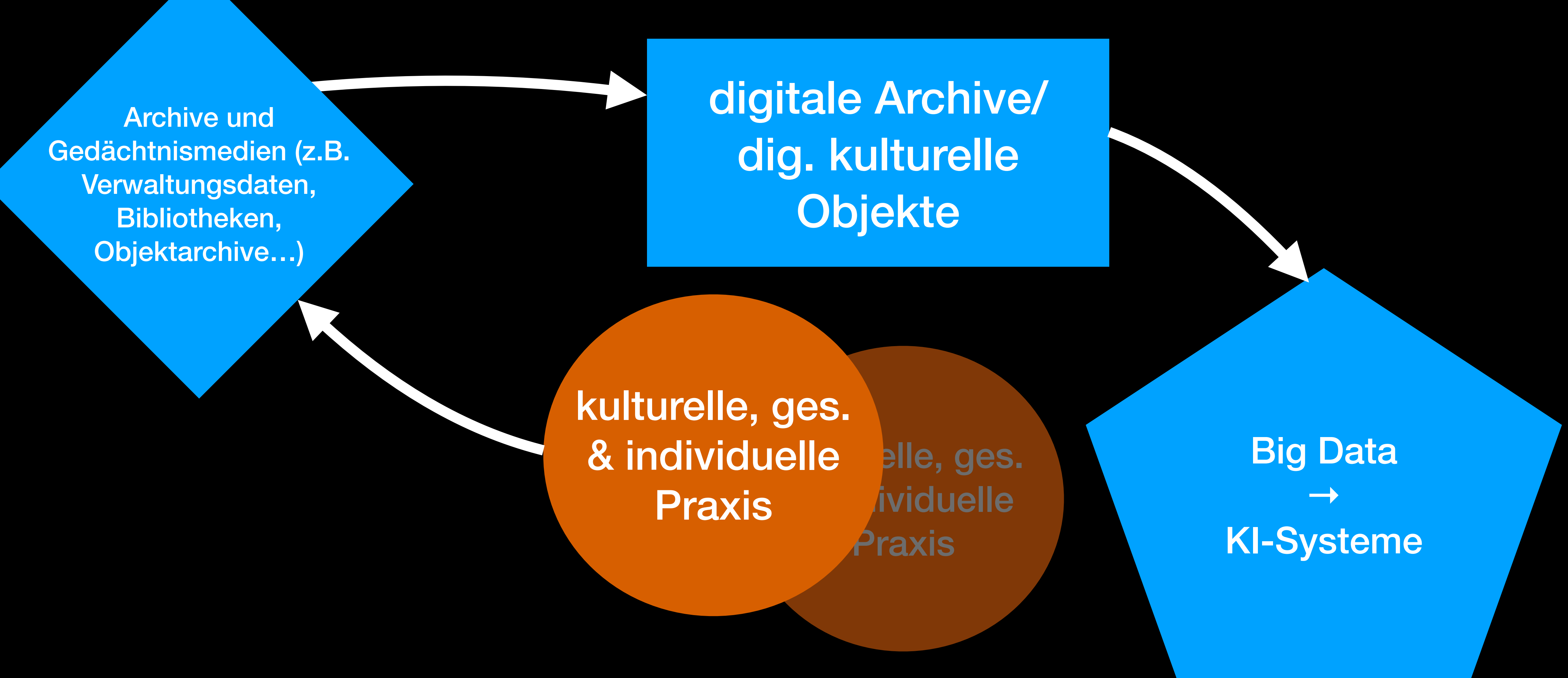
elle, ges.
individuelle
Praxis

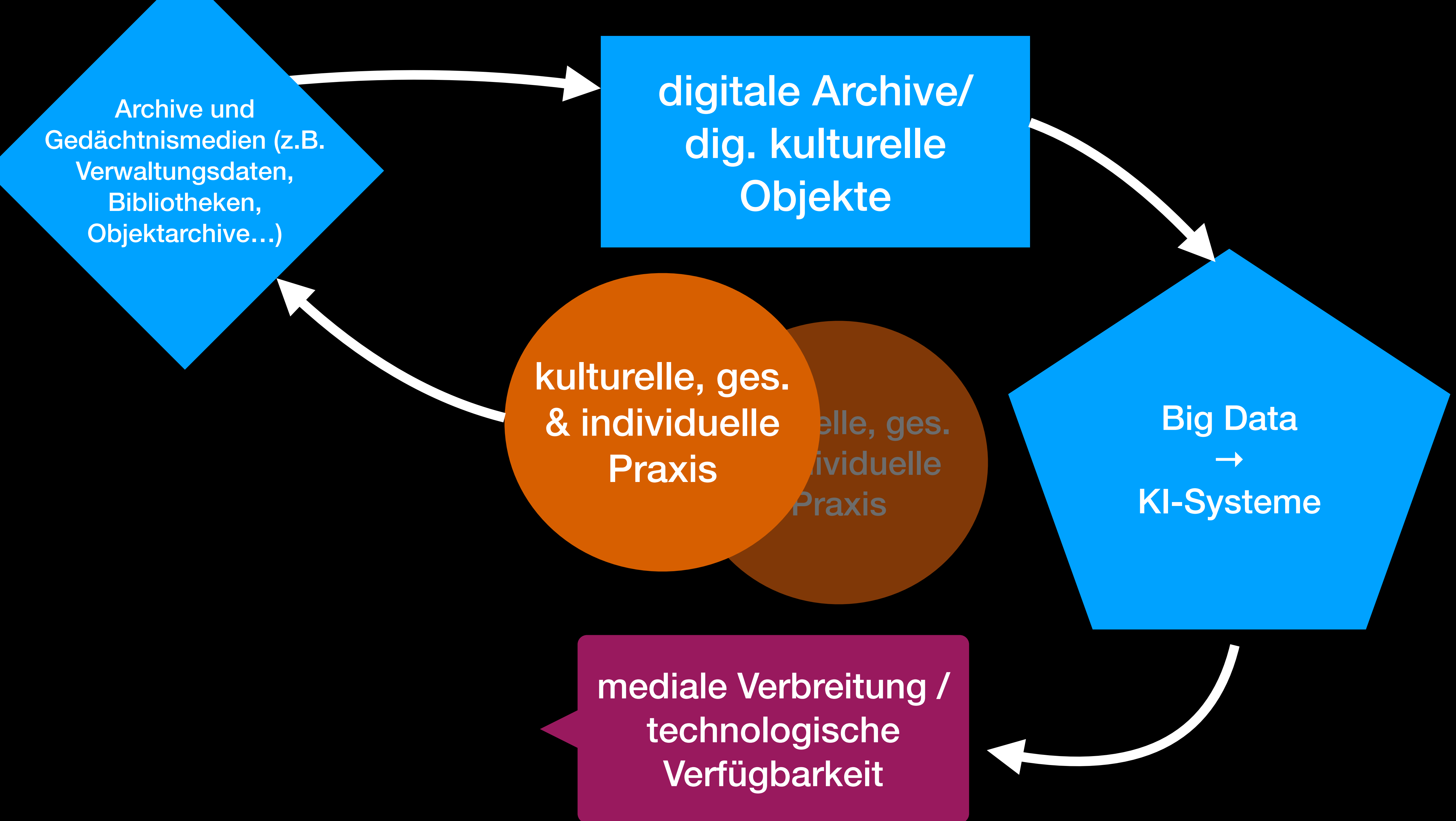
Archive und
Gedächtnismedien (z.B.
Verwaltungsdaten,
Bibliotheken,
Objektarchive...)

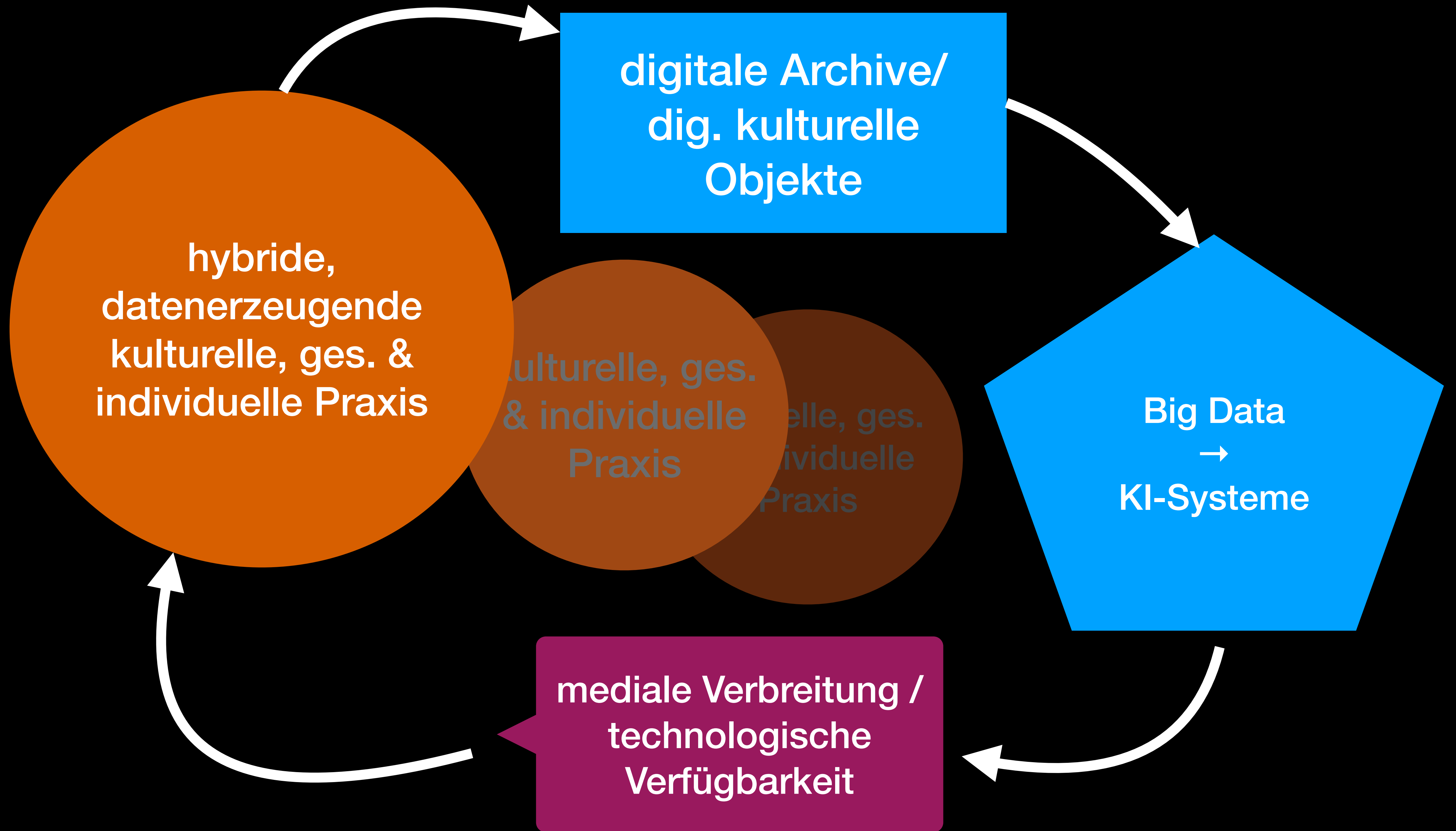
kulturelle, ges.
& individuelle
Praxis

kulturelle, ges.
& individuelle
Praxis









★ Member-only story

Ouroboros of AI: The peril of generative models feeding on their creations

Gilles de Peretti · [Follow](#)

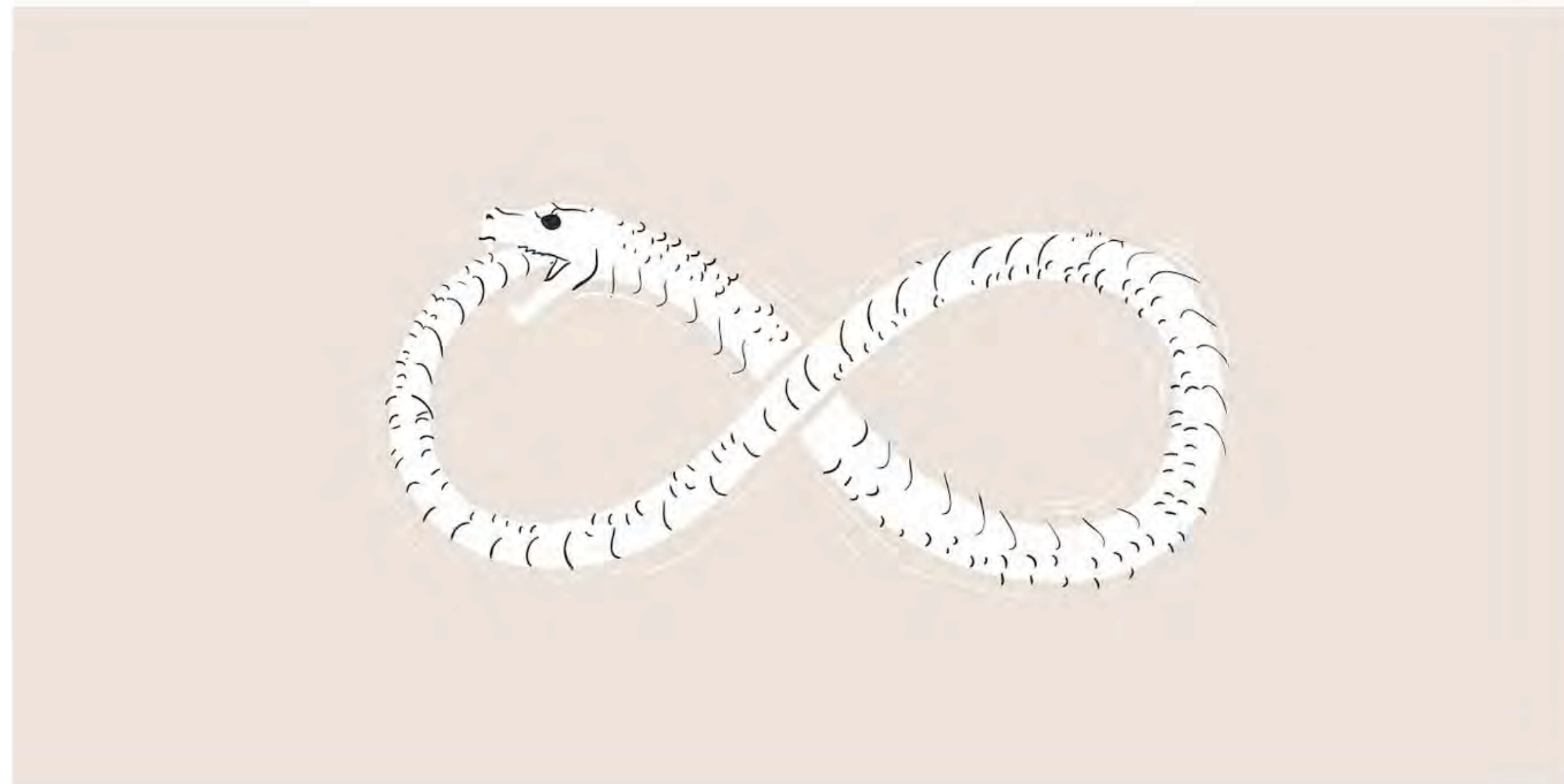
4 min read · Nov 19, 2023



84



1





Suchen...



NETZPOLITIK.ORG

♥ Spenden

Künstliche Intelligenz

Die schöne neue Welt der virtuellen Influencer

Eine neue Generation von Influencer:innen mischt weltweit die Szene auf. Die Sache ist nur: Es gibt sie nicht wirklich. Sie entstehen am Computer, haben zigtausende Follower und verwischen die Grenze zwischen Fiktion und Wirklichkeit.

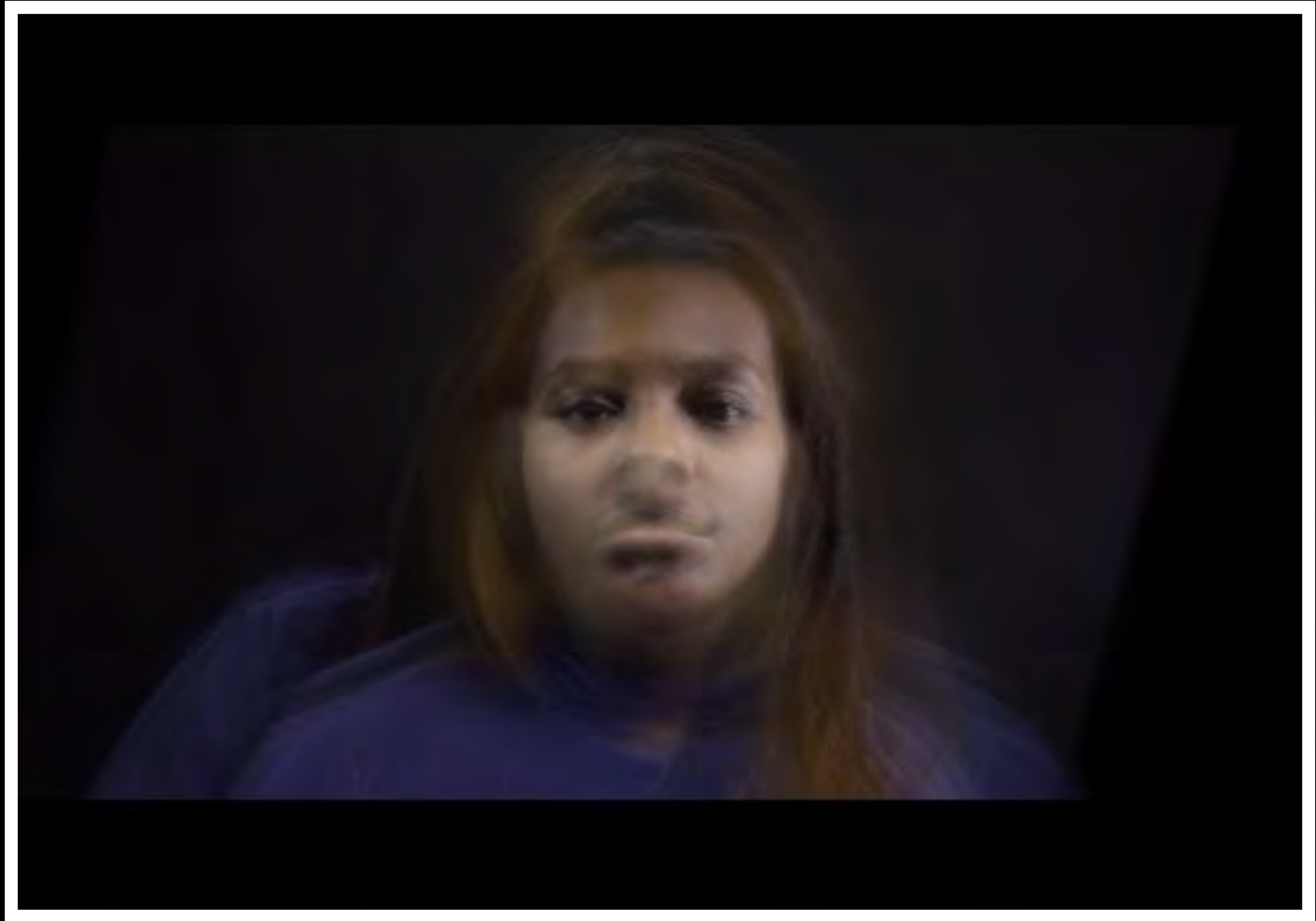
23.09.2023 um 06:59 Uhr - Markus Reuter - in Kultur - 14 Ergänzungen



Bildgeneratoren schaffen exakt die Ästhetik, die echte Influencer mühevoll inszenieren.

– Public Domain generiert mit Midjourney / Prompt: „woman on green palm tree at Tulum Beach,

artistic creation
as hybrid
aesthetic decision-
making



Screenshot: Holly Herndon & Jlin (feat. Spawn) – Godmother

Quelle: <https://youtu.be/sc9OjL6Mjgo> (Datum: 10.6.2020)



👤 : Andrés Mañón

I'm excited to finally share something I have been working on for the last year ✨ [Holly+](#)

I am releasing [Holly+](#) in collaboration with [Never Before Heard Sounds](#), the first tool of many to allow for others to make artwork with my voice, and will distribute ownership of my digital likeness through the creation of the Holly+ DAO 🧑🏻🧑🏻

My voice is precious to me! It is 1 of 1 🥰

Voice Models, in combination with machine learning technology, already allow for anyone to clone a voice to generate music and media, and the opportunities and complications inherent to these techniques will only intensify!

This development raises novel questions about voice ownership that I think can be addressed by DAO governance 🍷

Who am I?

I'm an artist and composer 🧑🏻 who has been working with machine learning for many years. My last album [PROTO\(4AD,2019\)](#) was the first to utilize singing neural networks and I completed my Doctorate at [Stanford's Center for Computer Research in Music and Acoustics](#), where my research focus was on the interplay between machine learning and the voice, and the implications of this technology for IP and vocal sovereignty 🧑🏻

Some AI models already know who I am! Here are some images spawned from my likeness using [OpenAI's CLIP model](#) 🧑🏻



Existenzialität vs. Nonexistenz

Kreativität vs. Rationalität

Fazit

Logik vs. Statistik

neues Wissen vs. Ouroboros-Effekt

Kontexte und Handlungsbedarfe

Künstliche Intelligenz als Wahrnehmungskrise: theoretische und ästhetische Herausforderungen

Rotary Club Erlangen, 7.3.2024

