

ML-Summit 2024

全球机器学习技术大会

Machine Learning Summit

—— 2024年4月25-26日 · 上海 ——

主办方: CSDn Boolan

合作伙伴:



元戎启行
DEEPROUTE.AI



☆ 2024技术会议—技术人的思想盛宴

全球机器学习 技术大会

ML-Summit 2024
www.ml-summit.org

9月5-6日 北京
12月5-6日 深圳

全球软件研发 技术大会

SDCon 2024
www.sdcon.com.cn

7月4-5日 北京
11月7-8日 上海

全球产品经理 大会

PM-summit 2024
www.pm-summit.org

8月8-9日 北京

全球C++及系 统软件技术大会

CPP-Summit 2024
www.cpp-summit.org

10月17-18日 深圳

☆ 媒体服务—多形态的**技术内容**

媒体服务



权威出版

《新程序员》

聚集全球技术大师、行业技术先锋，从前沿技术到行业先锋，为您带来全方位的深刻洞见和专业解读。

《中国开发者调查报告》

中国样本丰富的开发者调查报告，目前已成为完整准确了解中国开发者市场的重要参考资料。



扫码购买最新
版程序员杂志

全球机器学习技术大会

Daniel Povey

Kaldi之父, IEEE Fellow, 小米集团首席语音科学家



Daniel Povey 博士毕业于英国剑桥大学，2012年加入美国约翰霍普金斯大学，任语言和语音处理中心副教授。2019年10月加入小米，担任集团语音首席科学家。2022年12月，Daniel Povey 凭借在语音识别和声学建模方面的杰出贡献入选 IEEE Fellow。Daniel Povey 博士是著名开源语音识别工具 Kaldi 的提出者和主要维护者，被业界称为“Kaldi 之父”。Kaldi 项目在学术界和工业界都极具影响力，广泛应用于各研究机构和国内外知名公司，有力地推动了智能语音处理的产业化。

演讲主题：

The Current Moment in AI



The Current Moment in AI

Daniel Povey



Introduction

- I'm not going to talk much about specific technical trends
- I want to try to put the current excitement about LLMs in a longer historical context
- AI has a much longer journey to make
- It will evolve considerably, not just scale up.

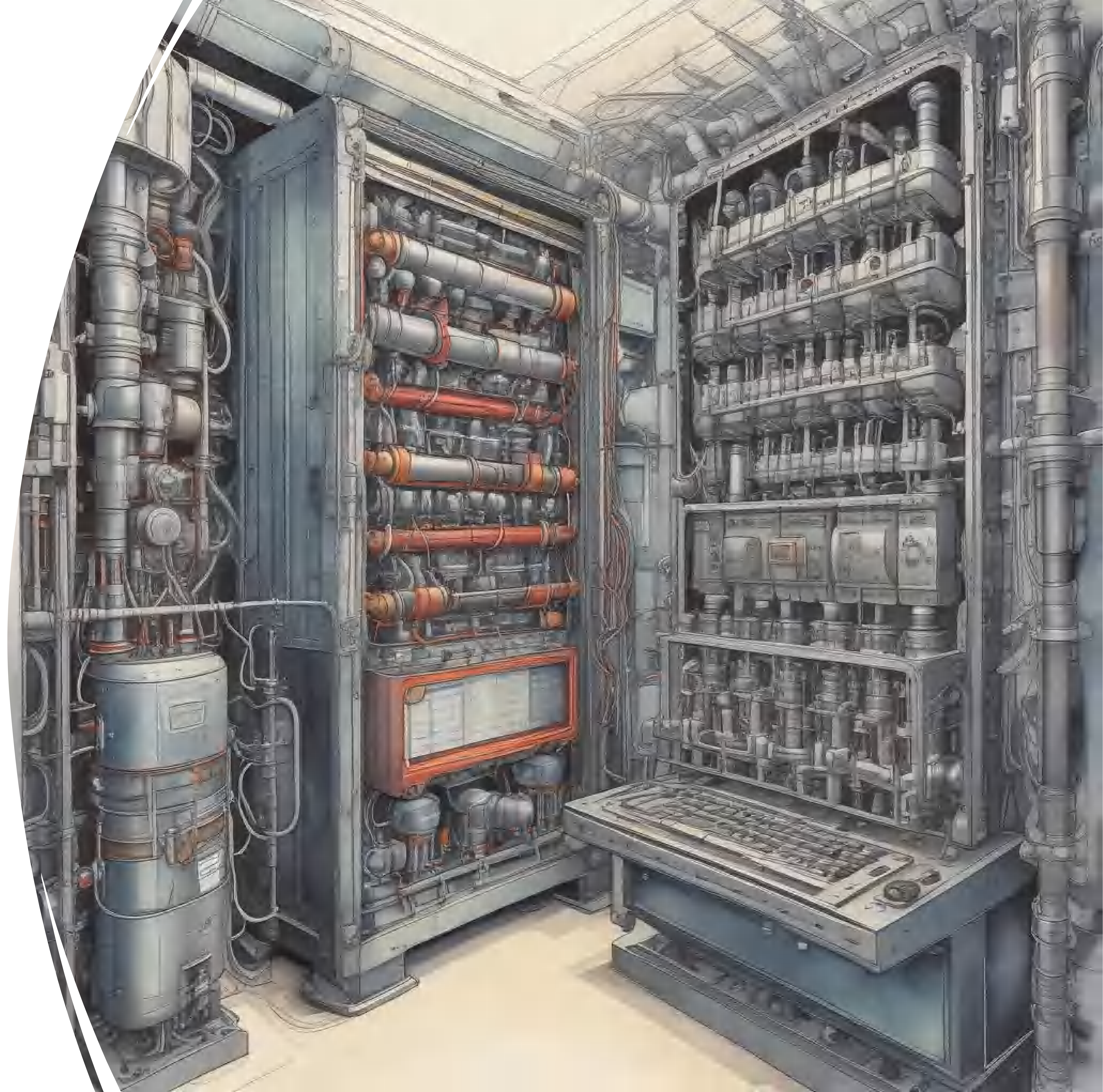


What have we learned from ChatGPT etc.?

- IMO we've learned:
 - We can get a model to learn a lot about the world with just a prediction loss
 - Scaling up the model and the training-data size is very helpful.
- Both of these things are good to know— but neither is very surprising.

Intelligent computers in Science Fiction

- In the sci-fi stories of Isaac Asimov, there was an enormous super-intelligent computer called Multivac
- Reminiscent of the real-life vacuum-tube-based computer “ENIAC”
- Reflects a tendency of our imagination to take existing tech and scale it up or extrapolate existing trends



“More” is not always necessary

- Human brain size has shrunk compared to Cro Magnon man (56,000 years ago)
- But we are probably capable of more things than we were at that time
- We may rely more on language, co-operation and communication than on detailed memory of our environment
- Humans now rely more *on knowing where to find information*, than remembering everything
 - This may be a sensible approach for AIs, as well



“Are we there yet?”

- On long car journeys, children often ask “are we there yet?”
- Most of the time, we are nowhere near there
- The development of AI will be a long process without an end
- The only end I think is likely is if human civilization regresses somehow



Large-scale experiments and lock-in

- By standardizing too early, it's possible to lock-in suboptimal designs
- The tendency to build very large models (often vanilla Transformers) and to fine-tune pretrained models tends to lock-in the architecture
- My own group's work includes new model topologies (Zipformer), layer types (BiasNorm), nonlinearities (SwooshR) and optimization algorithms (ScaledAdam)
- We are mostly doing medium-scale experiments, e.g. at most 10,000 hours of training data, not millions
 - No-one really does algorithm development at the million-hour scale



Our work: Zipformer

- Like a Transformer or Conformer for ASR, but with UNet-like structure: downsampling and upsampling
- Also a lot of other modifications
- State-of-the-art results when we compare with models trained & tested on the same datasets
- Also several times faster

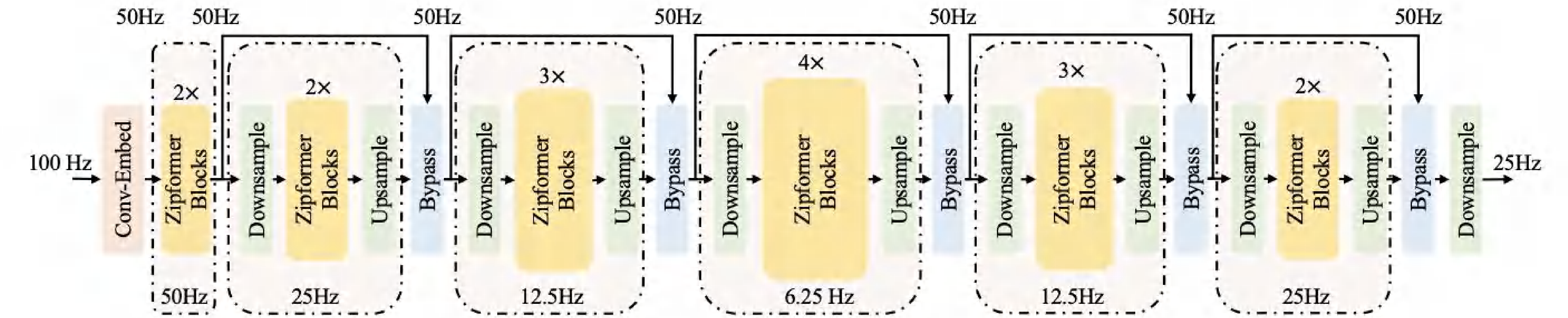


Figure 1: Overall architecture of Zipformer.

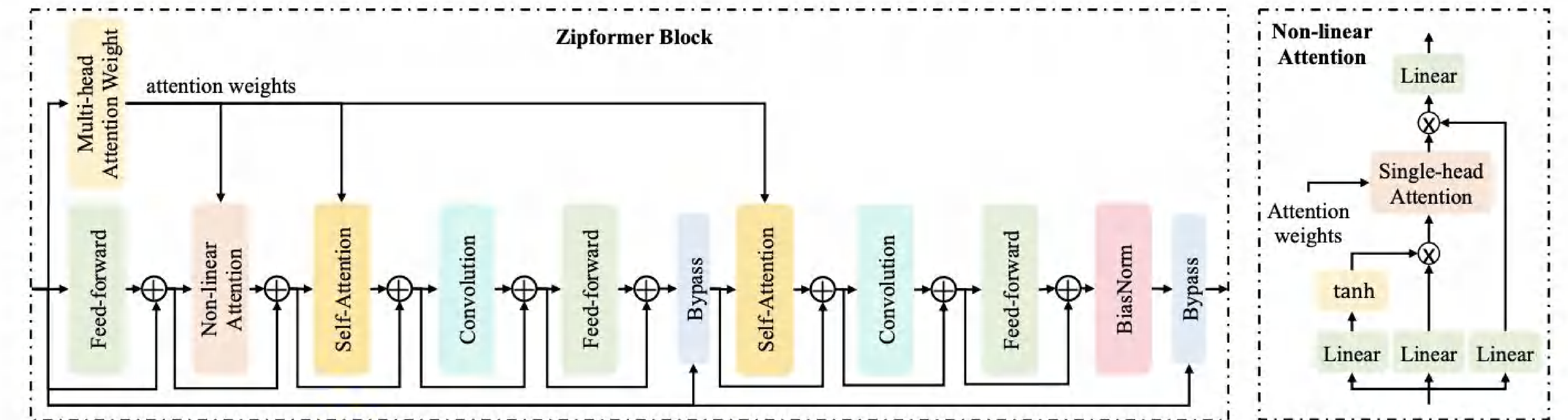


Figure 2: (Left): Zipformer block structure. (Right): Non-Linear Attention module structure.

Table 4: CER(%) comparison between different models on WenetSpeech dataset.

Model	Params (M)	Type	Dev	Test_Net	Test_Meeting
Conformer in ESPnet (Watanabe et al., 2018)	116.9	CTC/AED	9.70	8.90	15.90
Conformer in WeNet (Yao et al., 2021)	116.9	CTC/AED	8.88	9.70	15.59
Conformer-MoE(16e) (You et al., 2022)	425	CTC/AED, MoE	7.67	8.28	13.96
Conformer-MoE(32e) (You et al., 2022)	—	CTC/AED, MoE	7.49	7.99	13.69
Conformer-MoE(64e) (You et al., 2022)	—	CTC/AED, MoE	7.19	8.36	13.72
Zipformer-S	32.3	pruned transducer	7.96	8.6	13.97
Zipformer-M	75.9	pruned transducer	7.32	7.61	12.35
Zipformer-L	160.9	pruned transducer	7.29	7.24	12.06



Our work: ScaledAdam

- A version of the Adam optimizer that also learns parameter scales
- The main update of each tensor is scaled proportional to the current parameter norm
- More robust to changes in the model structure
- Makes it possible to get rid of most of the LayerNorm modules because learnable scales are “implicit”
- Faster optimization and improved results, in our setup at least.

Table 5: Ablation studies for *Zipformer-M*, including encoder structure, block structure, normalization layer, activation function and optimizer.

Ablation	Params (M)	test-clean (%)	test-other (%)
<i>Zipformer-M</i>	65.6	2.21	4.79
Encoder structure			
No temporal downsampling	94.2	2.23	5.09
Block structure			
Double Conformer-style blocks	73.9	2.18	4.95
No <i>NLA</i>	58.7	2.16	4.97
No <i>NLA</i> , no attention weights sharing	60.9	2.20	5.10
No <i>Bypass</i>	65.5	2.25	4.86
Normalization layer			
LayerNorm	65.6	2.29	4.97
Activation function			
Only <i>SwooshR</i>	65.5	2.32	5.21
Swish	65.5	2.27	5.37
Optimizer			
Adam	65.6	2.38	5.51

Our work: BiasNorm

- An improved version of LayerNorm
- By adding a bias before computing the norm, we “under-normalize”, meaning, some of the length information is retained
- Observation of model activations led us to conclude that the model does not “want” to fully normalize the activation vectors
- Also faster, because we take out the learnable bias and weight

Table 5: Ablation studies for *Zipformer-M*, including encoder structure, block structure, normalization layer, activation function and optimizer.

Ablation	Params (M)	test-clean (%)	test-other (%)
<i>Zipformer-M</i>	65.6	2.21	4.79
Encoder structure			
No temporal downsampling	94.2	2.23	5.09
Block structure			
Double Conformer-style blocks	73.9	2.18	4.95
No <i>NLA</i>	58.7	2.16	4.97
No <i>NLA</i> , no attention weights sharing	60.9	2.20	5.10
No <i>Bypass</i>	65.5	2.25	4.86
Normalization layer			
LayerNorm	65.6	2.29	4.97
Activation function			
Only <i>SwooshR</i>	65.5	2.32	5.21
Swish	65.5	2.27	5.37
Optimizer			
Adam	65.6	2.38	5.51

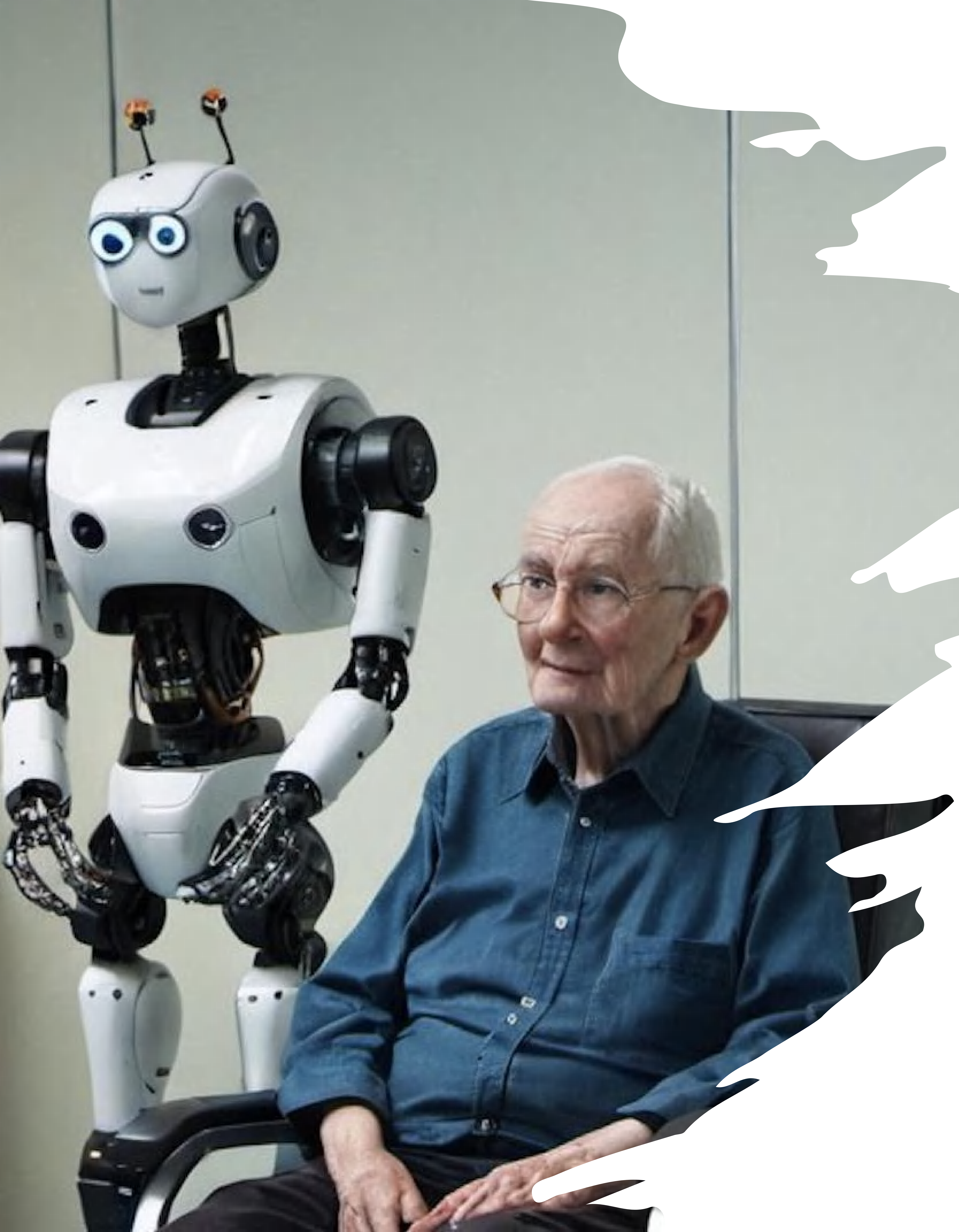
Do Androids Dream of Electric Sheep?

- This novel by Philip K. Dick asks whether androids are conscious.
- Movie adaptation: “Blade Runner”.
- I cannot answer RE consciousness, but the question of whether androids would dream is also interesting
- Dreaming has arisen separately in evolution
- Mammals, reptiles and birds all dream
- Octopuses also dream, although our most recent common ancestor had a too-simple nervous system to dream
- Perhaps dreaming is somehow a requirement for intelligence?
- Could it be part of some “generative” learning algorithm, perhaps related to a contrastive loss?



Rain Man

- In the movie “Rain Man”, an autistic character can say what day of the week any date falls on, but cannot navigate normal life.
- He can process information efficiently, but his motivation and selection of what tasks to focus on is not normal.
- In current LLM training, the model trains “equally” on all training tokens.
- Data may be from high quality or low quality sources
- I predict that future models will spend more time processing things that can help predict other things, especially across modalities (e.g. text vs. image).
 - We could make the model more “interested” in certain types of data or interactions
- Humans’ motivations are very complex and specific, dictated by our genes.
- We will probably want to build specific types of “motivation” into our AIs (whether by loss functions or by other means).

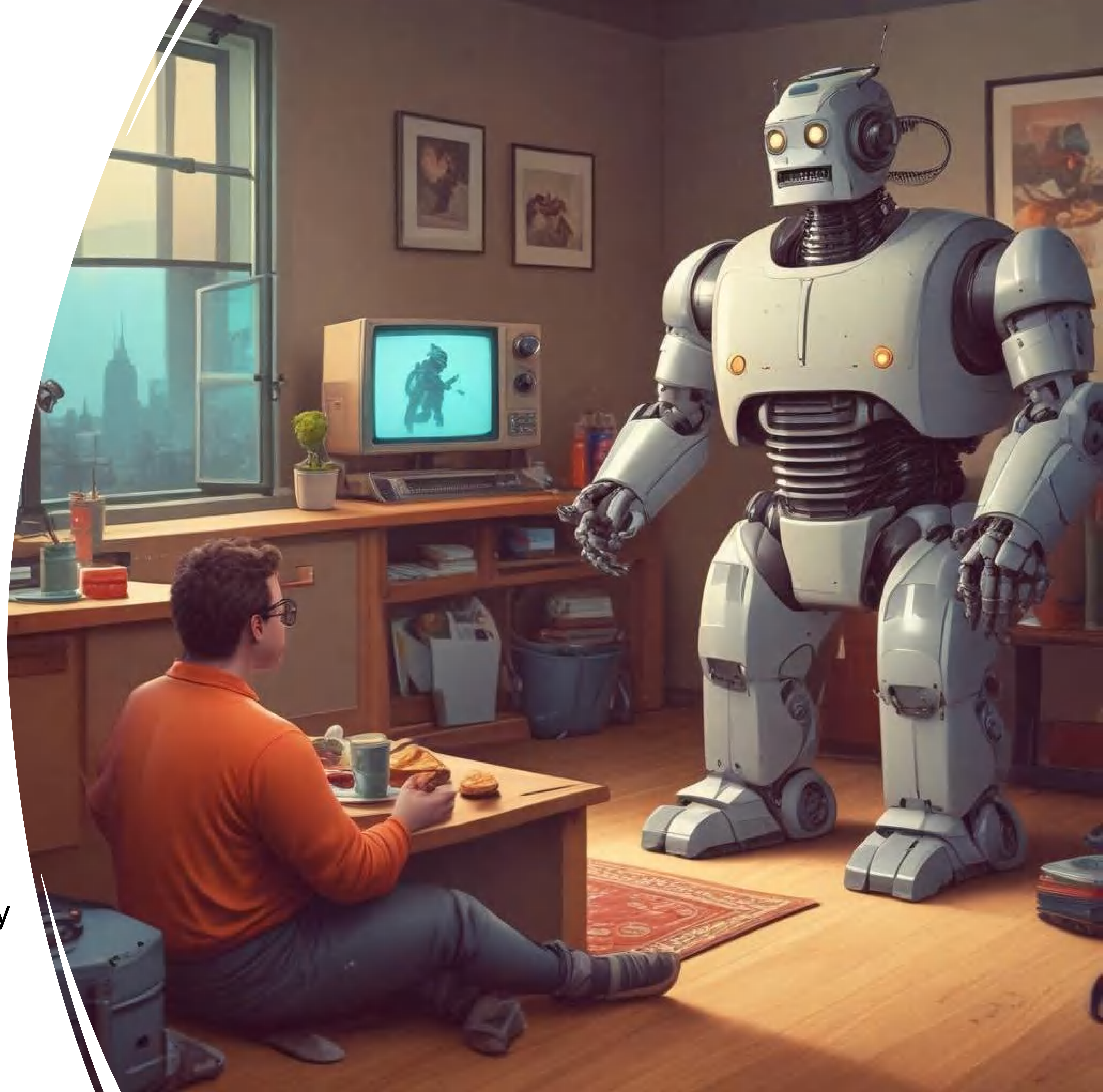


AI and Meaning

- We should not hope that AI will give us meaning or direction
- If anything, AI will make things seem more meaningless
 - E.g. AI beating us at chess makes us lose interest in chess
- People need answers to questions like: “Who am I? What is my purpose?”.
- AI cannot answer because it’s not a well-defined question.
- Ideas like “AI Accelerationism” are, I think, just a way to avoid addressing difficult questions.
- Concepts like AI solving all human problems are ridiculous because there aren’t “objectively right” solutions
- A dog with AI is still a dog

AI Utopia or Dystopia?

- People speak of an AI-enabled utopia...
- All the jobs are done by AI, and people are given an income for free so that they can consume
- This is a dystopia, not a utopia
- People derive meaning from their social role and their utility to, and connections with, others
- As useless atomized consumers, people would be very unhappy.
- Even if work were not necessary, we would probably have to invent it (and make it necessary somehow.)



AI's Midas Touch

- There is a legend of a King Midas that everything he touched turned to gold.
- This was a kind of curse, because his loved ones, his bed, his food and drink all turned to gold.
- I fear that generative AI could have a kind of Midas Touch for perceived value of media.
- Email has already become devalued by spam.
- Website content and search results are becoming dominated by AI-generated “fake” content.
- AI chatbots may cause people to distrust chatbots.
- AI phone calls may lead people to block unknown phone contacts.
- If AI becomes able to write fiction, it may make people lose interest in reading books by less-known authors.
- In science, the high volume of low-quality papers is already a problem and this will be made worse by AI.





The Library of Babel

- A story by Argentinian author Luis Jorge Borges about an infinite library
- It contains all possible books, including nonsense books
- The library is, of course, useless
- If there is too much low-quality data, we (and AIs) will have to be more discriminating in how we consume information.
 - E.g. check the reputation of the website, author or publisher before wasting our time by reading content
- I believe in future, AI training datasets will need good metadata and this should be included as part of the training algorithm
- Maybe even have the AI decide how to explore the data
- The space of all available information is, practically speaking, almost infinite. We cannot read randomly in an infinite library.

A man in a light blue shirt and dark trousers is running inside a large, circular, metallic tunnel. The tunnel has a complex, multi-layered structure with various pipes and supports. The man is captured in mid-stride, moving from left to right. The background is a bright, overcast sky.

AI and productivity of office workers

- It has been predicted that AI might increase the productivity of office workers or make many of them unnecessary
- Unlikely in the case of generative AI/LLMs
- In cases where the person mostly writes text for human consumption, one or both of the following is often true:
 - (A) The job exists because of regulation or licensing
 - (B) They are engaged in some kind of zero-sum competition
- Lawyers: (A) and (B)
- Advertising: (B)
- People who deal with environmental planning permission: (A) and (B)
- Human Resources: (A) and (B)
- Poets, authors, writers of music: (B)

AI and productivity in agriculture and mining

- There is huge potential to deploy AI in agriculture and mining
- These are rather unpredictable environments, so deploying “unintelligent” machines can be problematic
- They also operate at huge scale, which is ideal for AI
- AI makes it possible to use less environmentally harmful methods in agriculture and mining
- ... and less cruel methods in animal husbandry
- We need to develop AIs that are good at interacting with the physical world in an unstructured environment





AI and warfare

- AI and cheap drones are already revolutionizing warfare
- Makes killing much easier
- Enables highly asymmetrical wars, where the deaths are almost all on one side
- Reduces the advantage that defenders previously had (from fortifications, trenches, etc.)
- Also delocalizes war, because drones can strike deep in enemy territory
- If the primary victims of war are civilians, public anger may generate an impetus to continue a war... this might last until one side's ability to act is destroyed
- Guns made war more deadly because killing with guns requires less courage than with swords. Killer drones may be like this but even more, especially if fully AI (no human operator)

AI and education

- Education is a complex industry because there are different stakeholders with misaligned interests
- Families want their child to gain an advantageous social position
- Children want to have fun, enjoy their childhood and be popular
- Government or whoever controls the curriculum wants to:
 - Shape childrens' identity and opinions
 - Control what kinds of people get advantageous social positions
- Daytime childcare is also a key consideration for parents and the government.
- In general for mostly-zero-sum games, the application of AI won't fundamentally change things.
- In the past, the application of technology to education has not had much effect.



AI and communication

- AI will accelerate the trend of communication channels becoming clogged by spam and advertisements
- This will be particularly true of “open” protocols like telephone, text messages and e-mail.
- Even the internet outside of major social networks will be less and less trusted by users as AI content spreads.
- “Open” channels may become irrelevant as people move to “closed” social networks (e.g. WeChat/Weixin).
- These companies take on some of the responsibility to eject criminals/spammers.
- Different nations will continue to fight to control of what can be said and done on these platforms
- Most likely, major nations or blocs will have their own data ecosystems (otherwise, conflicting rules would be a problem.)



AI and consumer electronics: The spa vs. the bazaar

- In a spa, the experience is totally controlled by the spa owner, it is tranquil
- In a bazaar, anyone is free to enter, and you are constantly being approached
- AI will make the bazaar even more chaotic and the spa potentially more “feature-complete”
- In reality, consumer electronics always exist somewhere on this spectrum
- Users may want to be able to customize where they are on this scale at different times
- Humans may not perform at their best if they have constant mental stimulation
- Corporate reputation will be important for potential “spa” developers- do users trust you to be their filter?



AI and semi-skilled labor: transport, retail, food preparation

- Throughout history, there have been repeated panics about automation destroying jobs
- In the longer run, it never seems to have caused a problem, as new jobs appeared. But who knows.
- At the higher end, some work might be done by humans even though machines can do it, for social reasons.
- For international competitiveness, low-wage low-skill labor will be less and less of an advantage
- My concern is that we will rely more and more on systems that are centralized, complex and brittle
- For example, a hacker or simply a bug could cause all the automated cars of a particular provider to halt, stopping traffic in a whole country or worldwide
- Living in a highly-automated society could be like living in a tsunami-prone area



Conclusion

- AI has a lot further to develop– not just in size, but also in terms of the algorithms and the hardware used.
- In future I expect “cross-modal” AI, involving not just text but also images and other forms of data, will be important.
- Various types of “prediction loss” may still do a lot of heavy lifting...
- But I expect the AIs may become more selective about what data they choose to train on.
- We may endow AIs with “personalities” that dictate what kinds of information they prefer to train on, and in what way.
- Not all of the applications of AI will make our lives better.