ETH zürich

Distributed Computing Group

Seminar in Deep Neural Networks 2026

Frédéric Berdoz
17th of February 2026

Introduce yourself!

- Name
- Degree, Background in Machine Learning (theoretical and/or practical)
- What are your expectations for the seminar?
- What do you want to learn?

Supervisors

Frédéric

Joël

Andreas

Florian

Till

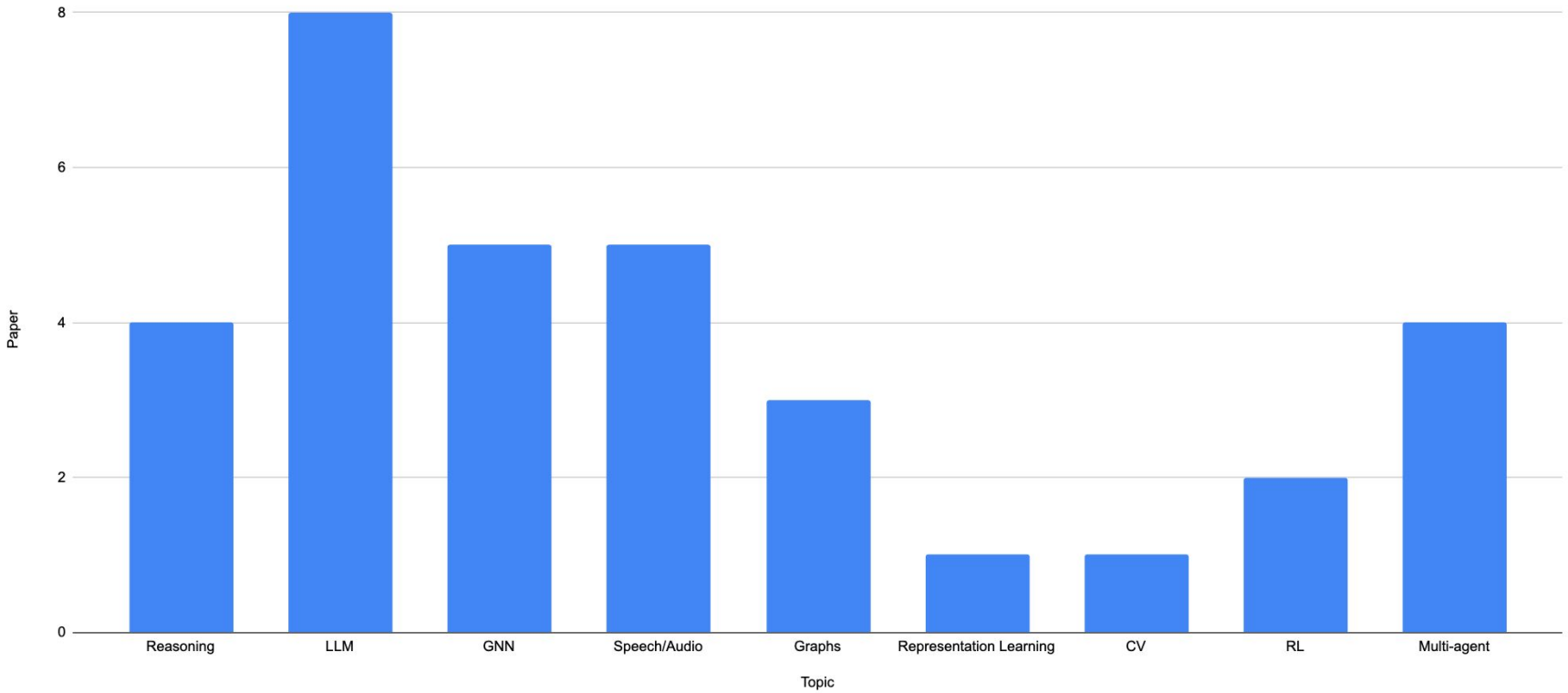
Luca

Saku

David

Antonio

Topic overview



Schedule

Before Easter break

Date	Presenter	Title	Mentor	Slides
February 17	Frédéric Berdoz	Introduction to Scientific Presentations	-	-
February 24	Jonathan Maillefaud	Generative Graph Pattern Machine	Saku Peltonen	-
	-	TBA	-	-
March 03	Tom Haidinger	Neural Message-Passing on Attention Graphs for Hallucination Detection	Florian Grötschla	-
	Yucheng Wang	MiMo-Audio: Audio Language Models are Few-Shot Learners	Luca Lanzendörfer	-
March 10	Frederico Aberle	Does Reinforcement Learning Really Incentivize Reasoning Capacity in LLMs Beyond the Base Model?	Frédéric Berdoz	-
	Daven Lim	Do Language Models Use Their Depth Efficiently?	Joël Mathys	-
March 17	Mara Godeanu	SAM Audio: Segment Anything in Audio	Florian Grötschla	-
	Klejdi Sevdari	VOICECRAFT: Zero-Shot Speech Editing and Text-to-Speech in the Wild	Luca Lanzendörfer	-
March 24	Rui Wang	WHALE: Towards Generalizable and Scalable World Models for Embodied Decision-making	David Jenny	-
	Maximilian Schlegel	Extending the Context of Pretrained LLMs by Dropping Their Positional Embeddings	Frédéric Berdoz	-
March 31	Ioana Gabor	The quest for the GRAph Level autoEncoder (GRALE)	Joël Mathys	-
	Peter Jolles	Graph Positional Encoding via Random Feature Propagation	Saku Peltonen	-

After Easter break

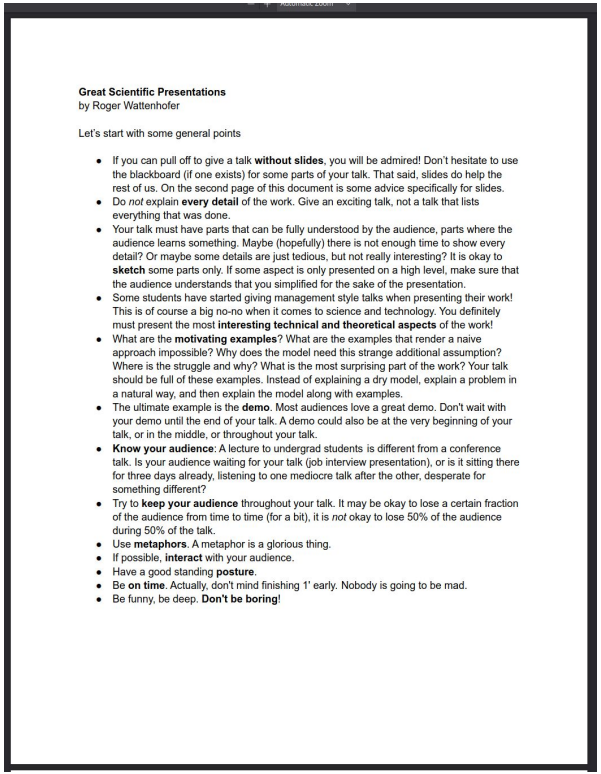
April 14	Prathamesh Tagore	Sign and Basis Invariant Networks for Spectral Graph Representation Learning	Saku Peltonen	-
	-	TBA	-	-
April 21	Benedict Armstrong	Energy-Based Transformers are Scalable Learners and Thinkers	Joël Mathys	-
	Yiming Wang	mHC: Manifold-Constrained Hyper-Connections	Frédéric Berdoz	-
April 28	-	TBA	-	-
	-	TBA	-	-
May 05	Felix Schatzl	DISCOVER: Automated Curricula for Sparse-Reward Reinforcement Learning	David Jenny	-
	Felix Quernheim	Generalizable Insights for Graph Transformers in Theory and Practice	Joël Mathys	-
May 12	Federico Lin	Differentiable Euler Characteristic Transforms for Shape Classification	Saku Peltonen	-
	Georgios Sideris	Collaborative Learning in the Jungle	Antonio Di Maio	-
May 19	Baraq Ben-Or Lipshitz	LeJEPA: Provable and Scalable Self-Supervised Learning Without the Heuristics	Andreas Plesner	-
	Samvit Mavinkurve	You Cannot Feed Two Birds with One Score: the Accuracy-Naturalness Tradeoff in Translation	Till Aczel	-
May 26	Lin Che	Overtrained Language Models Are Harder to Fine-Tune	Andreas Plesner	-
	Louis Barinka	What makes an image realistic?	Till Aczel	-

...

How to structure your talk

Introduction ~5min	Previous work ~10min	Contribution of the Paper ~15min
Why is this an interesting problem?	What has already been done to solve it?	What is newly proposed in this paper?

Presentation Style



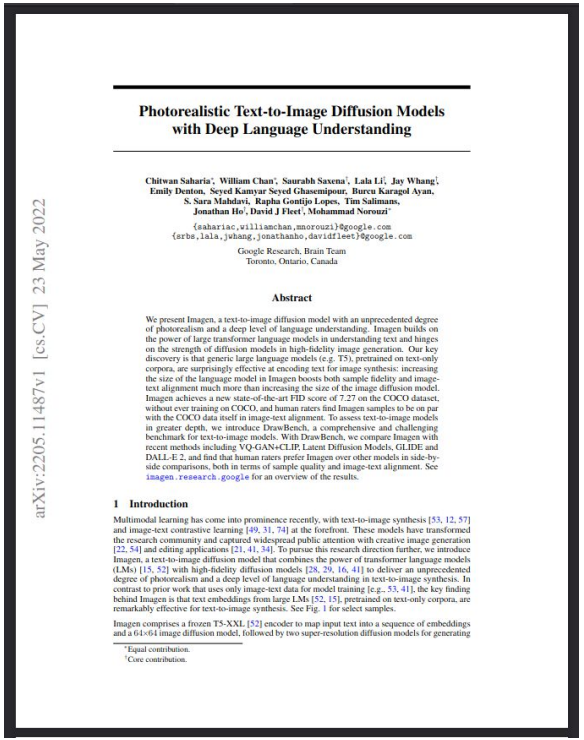
Great Scientific Presentations
by Roger Wattenhofer

Let's start with some general points

- If you can pull off to give a talk **without slides**, you will be admired! Don't hesitate to use the blackboard (if one exists) for some parts of your talk. That said, slides do help the rest of us. On the second page of this document is some advice specifically for slides.
- Do *not* explain **every detail** of the work. Give an exciting talk, not a talk that lists everything that was done.
- Your talk must have parts that can be fully understood by the audience, parts where the audience learns something. Maybe (hopefully) there is not enough time to show every detail? Or maybe some details are just tedious, but not really interesting? It is okay to **sketch** some parts only. If some aspect is only presented on a high level, make sure that the audience understands that you simplified for the sake of the presentation.
- Some students have started giving management style talks when presenting their work! This is of course a big no-no when it comes to science and technology. You definitely must present the most **interesting technical and theoretical aspects** of the work!
- What are the **motivating examples**? What are the examples that render a naive approach impossible? Why does the model need this strange additional assumption? Where is the struggle and why? What is the most surprising part of the work? Your talk should be full of these examples. Instead of explaining a dry model, explain a problem in a natural way, and then explain the model along with examples.
- The ultimate example is the **demo**. Most audiences love a great demo. Don't wait with your demo until the end of your talk. A demo could also be at the very beginning of your talk, or in the middle, or throughout your talk.
- **Know your audience**: A lecture to undergrad students is different from a conference talk. Is your audience waiting for your talk (job interview presentation), or is it sitting there for three days already, listening to one mediocre talk after the other, desperate for something different?
- Try to **keep your audience** throughout your talk. It may be okay to lose a certain fraction of the audience from time to time (for a bit), it is *not* okay to lose 50% of the audience during 50% of the talk.
- Use **metaphors**. A metaphor is a glorious thing.
- If possible, **interact** with your audience.
- Have a good standing **posture**.
- Be **on time**. Actually, don't mind finishing 1' early. Nobody is going to be mad.
- Be funny, be deep. **Don't be boring!**

https://disco.ethz.ch/courses/fs25/seminar/GreatScientificPresentations.pdf

How to approach your paper



arXiv:2205.11487v1 [cs.CV] 23 May 2022

Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding

Chitwan Saharia*, William Chan*, Saurabh Saxena†, Lala Li†, Jay Whang†, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S. Sara Mahdavi, Rapha Gontijo Lopes, Tim Salimans, Jonathan Ho†, David J Fleet†, Mohammad Norouzi*

{sahariac,williamchan,mnorouzi}@google.com
{srbs,lala,jwhang,jonathanho,davidfleet}@google.com

Google Research, Brain Team
Toronto, Ontario, Canada

Abstract

We present Imagen, a text-to-image diffusion model with an unprecedented degree of photorealism and a deep level of language understanding. Imagen builds on the power of large transformer language models in understanding text and hinges on the strength of diffusion models in high-fidelity image generation. Our key discovery is that generic large language models (e.g. T5), pretrained on text-only corpora, are surprisingly effective at encoding text for image synthesis: increasing the size of the language model in Imagen boosts both sample fidelity and image-text alignment much more than increasing the size of the image diffusion model. Imagen achieves a new state-of-the-art FID score of 7.27 on the COCO dataset, without ever training on COCO, and human raters find Imagen samples to be on par with the COCO data itself in image-text alignment. To assess text-to-image models in greater depth, we introduce DrawBench, a comprehensive and challenging benchmark for text-to-image models. With DrawBench, we compare Imagen with recent methods including VQ-GAN+CLIP, Latent Diffusion Models, GLIDE and DALL-E 2, and find that human raters prefer Imagen over other models in side-by-side comparisons, both in terms of sample quality and image-text alignment. See imagen.research.google for an overview of the results.

1 Introduction

Multimodal learning has come into prominence recently, with text-to-image synthesis [53, 12, 57] and image-text contrastive learning [49, 31, 74] at the forefront. These models have transformed the research community and captured widespread public attention with creative image generation [22, 54] and editing applications [21, 41, 34]. To pursue this research direction further, we introduce Imagen, a text-to-image diffusion model that combines the power of transformer language models (LMs) [15, 52] with high-fidelity diffusion models [28, 29, 16, 41] to deliver an unprecedented degree of photorealism and a deep level of language understanding in text-to-image synthesis. In contrast to prior work that uses only image-text data for model training [e.g., 53, 41], the key finding behind Imagen is that text embeddings from large LMs [52, 15], pretrained on text-only corpora, are remarkably effective for text-to-image synthesis. See Fig. 1 for select samples.

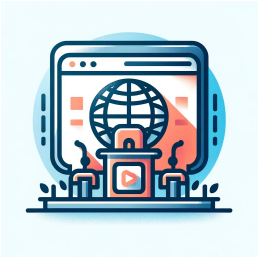
Imagen comprises a frozen T5-XXL [52] encoder to map input text into a sequence of embeddings and a 64×64 image diffusion model, followed by two super-resolution diffusion models for generating

*Equal contribution.
†Core contribution.

OpenReview.net

ICLR

NEURAL INFORMATION PROCESSING SYSTEMS

Admin stuff

- **Attendance:** Mandatory for all sessions
- **Active participation:** Ask questions, this will be graded
- **Presentation setup:** Use your own laptop
- **Recording:** All presentations will be recorded for final grading
- **Submission:**
  - Slides due the evening before your presentation
  - Mock presentation slides must also be submitted to your supervisor
- **Grading:**
  - Participation contributes to your grade
  - Final grades released after the seminar
- **Resources:** Check the website for updated information