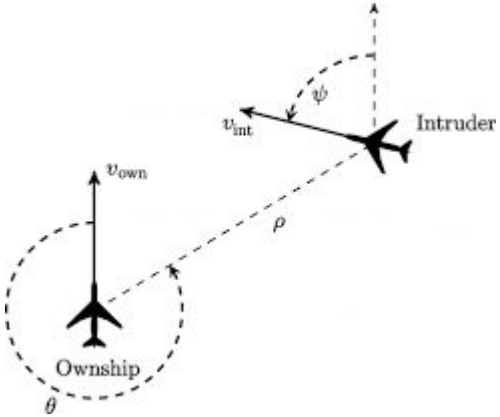
Recent Advances in the Verification of Neural Networks

first meeting

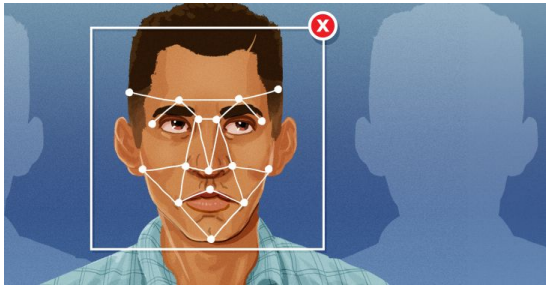
Max Prokop ● Stefanie Mohr ● Sabine Rieder ● Jan Kretinsky

Why?

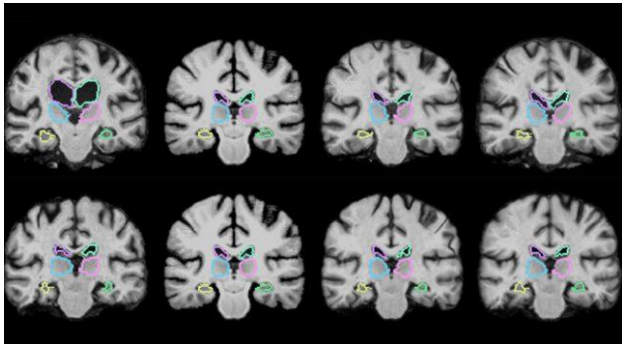
https://netwalk.de/de/service-2/



https://www.researchgate.net/figure/The-ACAS-Xu-System-adapted-from-Katz-et-al-2017_fig1_334695061



https://mashable.com/article/amazon-rekognition-software-sense-fear



https://scitechdaily.com/machine-learning-algorithm-compares-3d-scans-up-to-1000-times-faster/

DeepL

deepl.com

Outline

Motivation and Structure

Grading and Expectations

- Written Work, Presentation(s), Discussion

Discussion

- Suggested topics
- Schedule

Outline

Motivation and Structure

Grading and Expectations

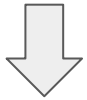
- Written Work, Presentation(s), Discussion

Discussion

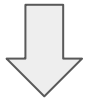
- Suggested topics
- Schedule

Motivation

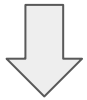
Learn about: (1) scientific writing (2) scientific presentations (3) your topic



Good & interesting presentations



Deepening discussions

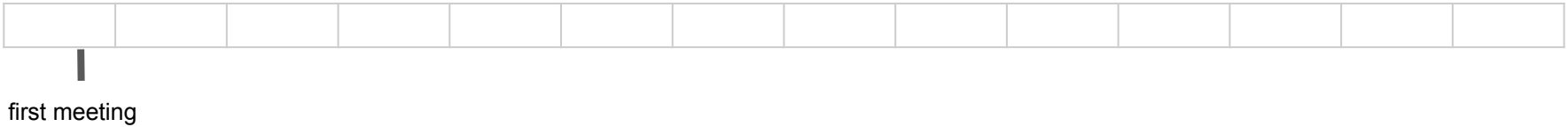


Find connections between topics and interesting research questions

High level idea

- read a scientific paper
- understand the paper
- write a short summary of it (2-4 pages)
- present it

Structure

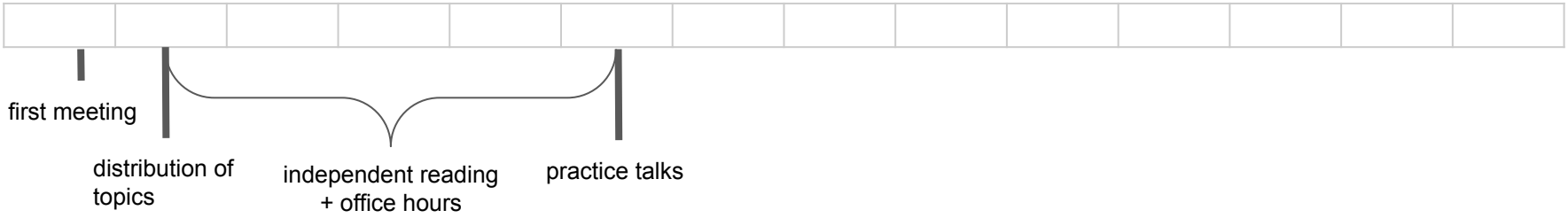


Structure

- Read the abstracts of the papers
- Skim through them
- Decide which interests you most

https://web.stanford.edu/class/ee384m/Handouts/HowtoReadPaper.pdf

Structure


first meeting
distribution of topics
independent reading + office hours
practice talks

Structure


first meeting
distribution of topics
independent reading + office hours
practice talks
independent reading + office hours
abstract submission

Structure


first meeting
distribution of topics
independent reading + office hours
practice talks
independent reading + office hours
abstract submission
final presentation

Outline

Motivation and Structure

Grading and Expectations

- Written Work, Presentation(s), Discussion

Discussion

- Suggested topics
- Schedule

Grading and Expectations

We expect you to attend all presentations, including the ones **after** your talk (not attending will result in failing the seminar)

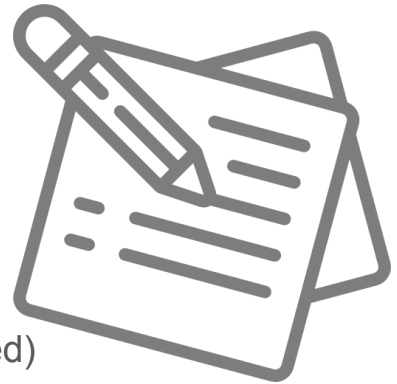
Your final grade will be determined from

 Written work (40%)

 Presentation (40%)

 Discussion + Chairing (20%)

Written Work



- Send **one week before** presentation
- 2-4 pages extended abstract (Fontsize 11, Latex template will be provided)
- Basis for the talk
- Basis for the listener to prepare questions
- Suggested structure
  - Motivation
  - Problem statement
  - Preliminaries
  - Outlook to the talk

Written Work - Grading Criteria

- Subject knowledge
- Subject correctness
- Structure
- Story
- Visuals
- Mechanics
- Scientific conduct

Presentation

- 20min oral presentation
- supported by slides (or whiteboard if you dare)
- present your paper in a way that the audience understand it (be brave to leave out formulae!)

Presentation - Grading Criteria

- Content
- Subject correctness
- Structure
- Story
- Visuals
- Mechanics
- Scientific conduct
- Timing
- Verbal skills (enthusiasm, speaking skills)
- Nonverbal skills (eye contact, body language, poise)

Discussion

- As participant
  - Send two questions per talk (based on abstract)
  - Participate in the discussion
- As chair
  - Introducing speaker
  - Timing of the talk
  - Leading discussion

Discussion - Grading Criteria

- Attendance
- Questions
- Participant - Wortbeitrag
- Chairing

Practice Talk?

- Only for you to practice!
- 5min
- Use it
  - a) as a teaser (make people interested)
  - b) just as a practice to get feedback
- Will not be graded, but use it for yourself to deliver the best in the final presentation!

Outline

Motivation and Structure

Grading and Expectations

- Written Work, Presentation(s), Discussion

Discussion

- Suggested topics
- Schedule

Suggested Topics

Please send a list of **6** papers in the order of your preferences to us by **EOD, Tue, 24.10.**

Steffi:

- An Abstraction-Refinement Approach to Verifying Convolutional Neural Networks (Ostrovsky, Barrett, Katz, ATVA22)
- Formal verification of neural agents in non-deterministic environments (Akintunde, Botoeva, Kouvaros, Lomuscio, 22)
- Linearity Grafting: Relaxed Neuron Pruning Helps Certifiable Robustness (T Chen, H Zhang, Z Zhang, S Chang, S Liu, P Chen, Z Wang)
- An Abstraction-Based Framework for Neural Network Verification (Yizhak Yisrael Elboher, Justin Gottschlich, Guy Katz, CAV 2020)

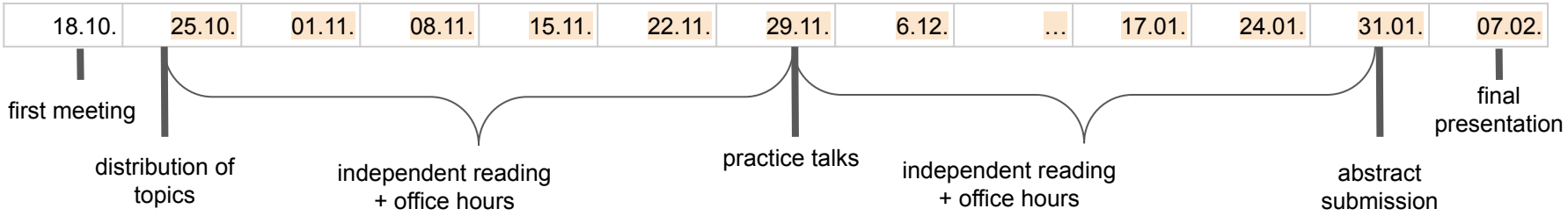
Sabine:

- Verifying learning-augmented systems (Eliyahu, Kazak, Katz, Schapira, ACM 2021)
- CLEVEREST: Accelerating CEGAR-based Neural Network Verification via Adversarial Attacks (Zhao et al. SAS 2022)
- Verification of Deep Convolutional Neural Networks Using ImageStars (Tran, Xiang, Johnson, CAV 2020)
- Improving Neural Network Verification through Spurious Region Guided Refinement (Yang et al., TACAS 2021)
- Complete Verification via Multi-Neuron Relaxation Guided Branch and Bound (Ferrari, Müller, Jovanovic, Vechev, ICLR 2022)
- Neuro-Symbolic Verification of Deep Neural Networks (Xie, Kersting, Neider, IJCAI 2022)

Max:

- Safe Reinforcement Learning via Shielding (Mohammed Alshiekh, Roderick Bloem, Rüdiger Ehlers, Bettina Könighofer, Scott Niekum, Ufuk Topcu)
- Deep Statistical Model Checking (Timo P. Gros, Holger Hermanns, Jörg Hoffmann, Michaela Klauck, and Marcel Steinmetz)
- Shielded Reinforcement Learning for Hybrid Systems (Asger Horn Brorholt, Peter Gjøl Jensen, Kim Guldstrand Larsen, Florian Lorber, and Christian Schilling)

Schedule Suggestion



Office hours:
Wednesday 11-13
send a mail with the questions at least 24h in advance