8. Limitations and future vision

Generative Music AI

THE SOUND OF AI

upf. Universitat Pompeu Fabra Barcelona

MTG Music Technology Group

Overview

- Limitations of state-of-the-art systems
- Limitations in research procedure
- New directions

Text-to-music (e.g., MusicLM)

- Long-term structure

Text-to-music (e.g., MusicLM)

- Long-term structure
- Audio fidelity

Text-to-music (e.g., MusicLM)

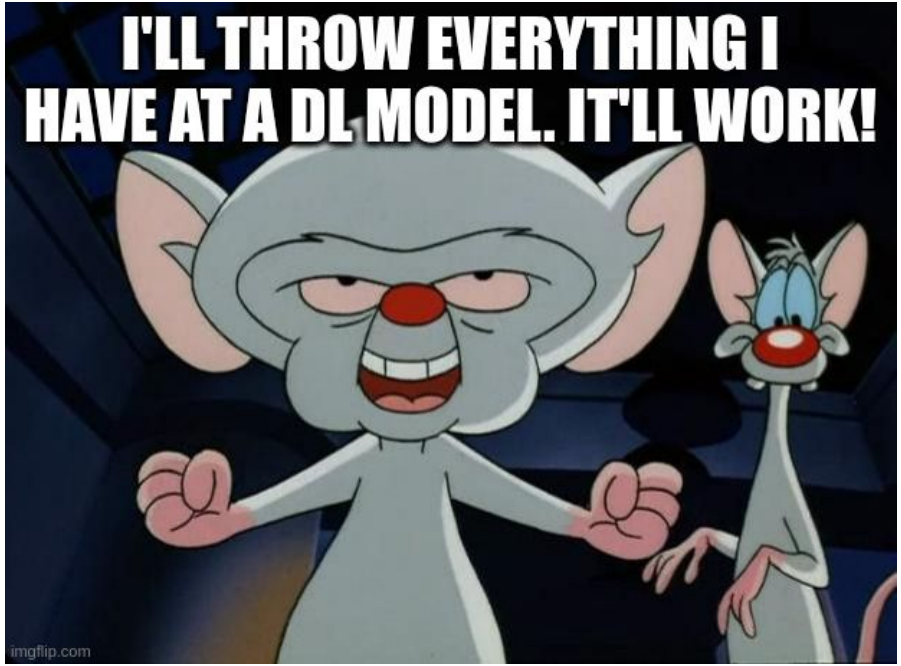
- Long-term structure
- Audio fidelity
- Semantic mapping

Text-to-music (e.g., MusicLM)

- Long-term structure
- Audio fidelity
- Semantic mapping
- Minimal creative control


GEN MUS
RESEARCHER
DEEP LEARNING
MUSIC
KNOWLEDGE
imgflip.com

The curse of Deep Learning



Problems with DL models

- Music is highly dimensional

Problems with DL models

- Music is highly dimensional
- Network can’t learn all dimensions

Problems with DL models

- Music is highly dimensional
- Network can’t learn all dimensions
- No musical knowledge

Problems with DL models

- Music is highly dimensional
- Network can’t learn all dimensions
- No musical knowledge
- Massive datasets

Problems with DL models

- Music is highly dimensional
- Network can’t learn all dimensions
- No musical knowledge
- Massive datasets
- Lack of music coherence

Problems with DL models

- Music is highly dimensional
- Network can’t learn all dimensions
- No musical knowledge
- Massive datasets
- Lack of music coherence
- Black box -> difficult to steer

Problems with DL models

Yann LeCun
@ylecun

On the highway towards Human-Level AI, Large Language Model is an off-ramp.

10:39 AM · Feb 4, 2023 · **1.5M** Views

340 Retweets **168** Quotes **3,046** Likes **419** Bookmarks

A Path Towards Autonomous Machine Intelligence
Version 0.9.2, 2022-06-27

Yann LeCun
Courant Institute of Mathematical Sciences, New York University yann@cs.nyu.edu
Meta - Fundamental AI Research yann@fb.com

June 27, 2022

Abstract

How could machines learn as efficiently as humans and animals? How could machines learn to reason and plan? How could machines learn representations of percepts and action plans at multiple levels of abstraction, enabling them to reason, predict, and plan at multiple time horizons? This position paper proposes an architecture and training paradigms with which to construct autonomous intelligent agents. It combines concepts such as configurable predictive world model, behavior driven through intrinsic motivation, and hierarchical joint embedding architectures trained with self-supervised learning.

Keywords: Artificial Intelligence, Machine Common Sense, Cognitive Architecture, Deep Learning, Self-Supervised Learning, Energy-Based Model, World Models, Joint Embedding Architecture, Intrinsic Motivation.

Solving the curse of DL

- Hybrid systems

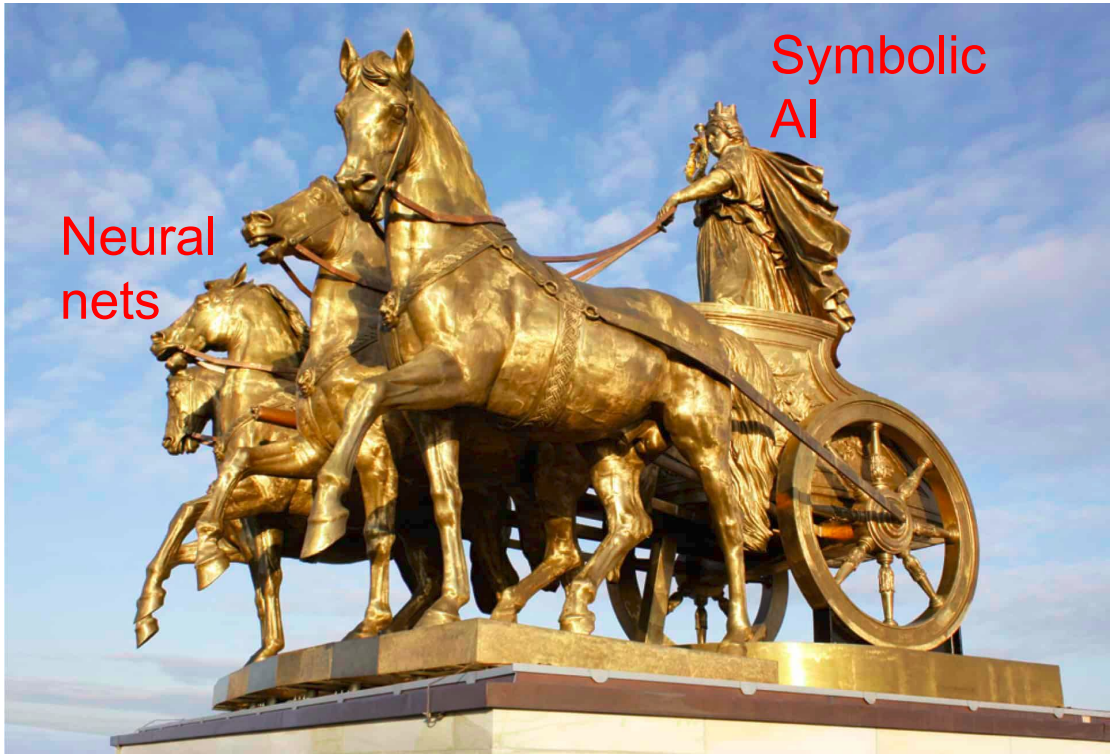
Solving the curse of DL

- Hybrid systems
- Merge DL and symbolic AI
  (Neuro-Symbolic Integration)

Solving the curse of DL

- Hybrid systems
- Merge DL and symbolic AI (Neuro-Symbolic Integration)
- Orientate the wilderness of generative DL models with reasoning on music knowledge

Solving the curse of DL



Music representation

- Audio is too complex
- Symbolic is too simple
- No representation captures all music details efficiently

Music representation

Music representation: Way forward

- Hybrid symbolic + audio representations

Music representation: Way forward

- Hybrid symbolic + audio representations
- Embeddings (symbolic + audio + context)

Music representation: Way forward

- Hybrid symbolic + audio representations
- Embeddings (symbolic + audio + context)
- Custom representations


SOMEBODY SAID
USER INTERFACES?
imgflip.com

User interfaces: Way forward

- UI as important as model
- User at the centre
- User evaluation

Problems with research procedure

Problems with research procedure

- No code / no model released -> not reproducible

Problems with research procedure

- No code / no model released -> not reproducible
- Matrioska systems based on proprietary models

Problems with research procedure

- No code / no model released -> not reproducible
- Matrioska systems based on proprietary models
- Smaller players can’t compete -> AI oligopoly (Google, OpenAI)?

Research procedure: Way forward

- Open source (code, models, datasets)

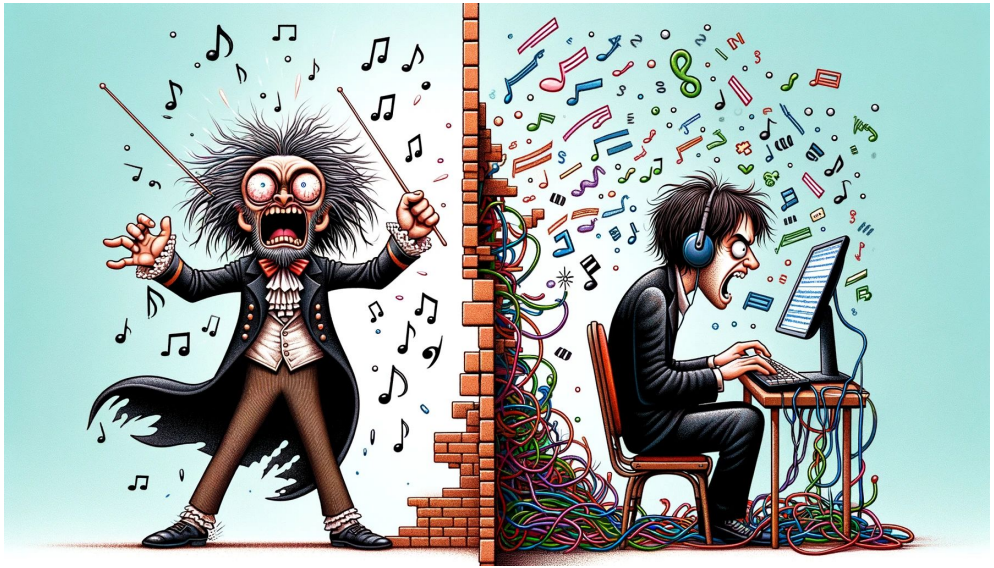
Research procedure: Way forward

- Open source (code, models, datasets)
- Reject papers without code / models

Research procedure: Way forward

- Open source (code, models, datasets)
- Reject papers without code / models
- Small models
  - Smart >> big
  - Distillation, pruning, …
  - Data quality

Problems with research procedure



Problems with research procedure

Key takeaways

- Don’t fall for the curse of DL

Key takeaways

- Don’t fall for the curse of DL
- Develop hybrid systems (DL + symbolic AI)

Key takeaways

- Don’t fall for the curse of DL
- Develop hybrid systems (DL + symbolic AI)
- Symbolic / audio representations have limitations

Key takeaways

- Don’t fall for the curse of DL
- Develop hybrid systems (DL + symbolic AI)
- Symbolic / audio representations have limitations
- UI is overlooked

Key takeaways

- Don’t fall for the curse of DL
- Develop hybrid systems (DL + symbolic AI)
- Symbolic / audio representations have limitations
- UI is overlooked
- Big tech oligopoly

Key takeaways

- Don’t fall for the curse of DL
- Develop hybrid systems (DL + symbolic AI)
- Symbolic / audio representations have limitations
- UI is overlooked
- Big tech oligopoly
- Open source + small models

Key takeaways

- Don’t fall for the curse of DL
- Develop hybrid systems (DL + symbolic AI)
- Symbolic / audio representations have limitations
- UI is overlooked
- Big tech oligopoly
- Open source + small models
- Engineers + music experts collaboration is key

What next?

Generative grammars