Enhancing Education with AI

Mihai Dascalu, NUST Politehnica Bucharest

Universitatea Națională de Știință și Tehnologie POLITEHNICA București

About Me

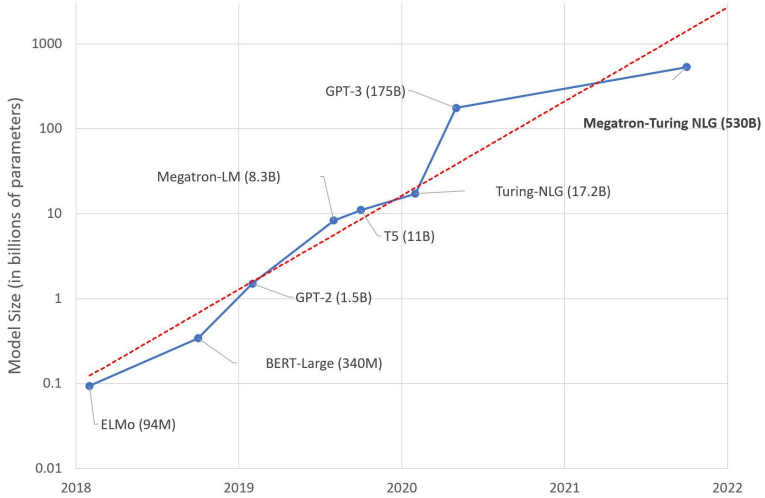
- Prof in CS @ Politehnica Bucharest
- Double PhD in CS & Educational Sciences
- 15+ years of building AI solutions for education
- Focus on comprehension assessment
- 300+ papers
    - 25+ papers in AIED & 15+ in ITS (Top conferences)

Emergence of Gen AI

The facts

- Spectacular progress [1]
  - NLP landscape reshaped by the Transformer
- Funding surge - 2023: $50 billion [2]
- “The ChatGPT Revolution”

A new Moore’s Law



The cons

- Ethical concerns
  - Authenticity ?
  - Deepfakes (faceswap, stable diffusion, GANs)
- Computational and environmental costs:
  - HW race - Meta target 340k+ H100 [3]
- Job displacement
- Security risks - prompt injection

[1] https://techjury.net/blog/ai-statistics
[2] https://news.crunchbase.com/ai/hot-startups-2023-openai-anthropic-forecast-2024/
[3] https://www.theverge.com/2024/1/18/24042354/mark-zuckerberg-meta-agi-reorg-interview

AI in Education - A Rapidly Evolving Field

- Personalized learning
  - Adaptive learning & recommendations
  - Identify learning gaps
- Task automation
  - Grading essay & multi-choice tests
  - Personalized learning paths
- Enhanced engagement
  - Gamification
  - Early identification of at-risk students
- Intelligent tutoring and feedback
- Immersive learning experiences - mixed realities

- Concerns on data privacy & security
- Mitigating biases & inequalities
- Dependence on technology detrimental to interpersonal skills
- Redefining the teacher role
- Quality and effectiveness

Our Recent Work in AIED

Scoring Systems Across the Years

Small Datasets >> Large Corpora

- Initially: <1k summaries on ~30 texts
- 5k+ summaries on 100+ texts
  - 7 dimensions: cohesion, objective language, paraphrasing, language beyond source text, summary length, details, main points
- Commonlit Kaggle summary scoring competition - 24k summaries
  - Wording & content dimensions

Models

- Wide range of linguistic features integrated in our open-source framework ReaderBench
  - Surface, Syntax, Semantics, Word
  - Classic Machine Learning models on top
- Transformer-based Encoders (including Longformers for longer contexts)
- Multi-task learning
- AutoML pipeline

Summary Scoring Cont.

Results

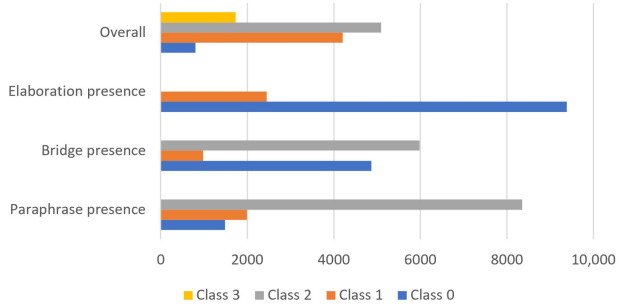
- MTL surpassed single task models & models built using ReaderBench linguistic features
- Average normalized MAE of .13 & R^2 of .55
- Predict human rating of a summary with an average deviation of 13%

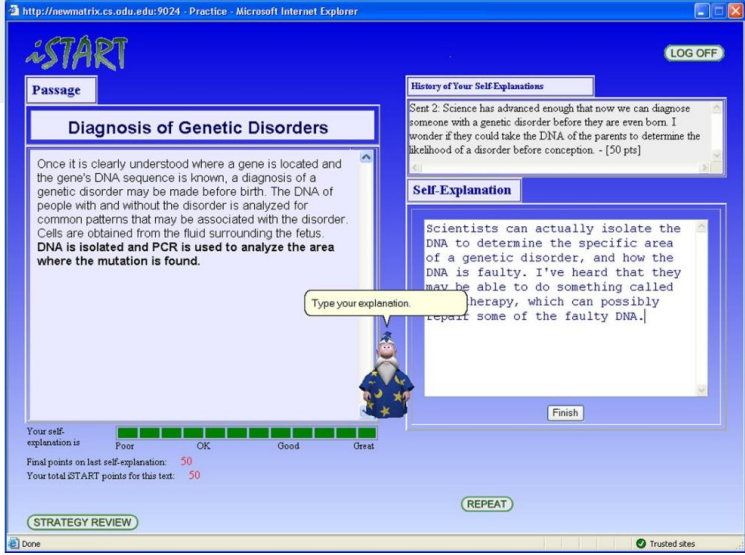
Takeaways

- Robust models
- Explainability remains a major concern
- Aims:
  - Generating exemplars
  - Generating personalized feedback with recommendations

Specific Instructions Self-Explanation

- 3 datasets (N = 11,833) with SEs annotated on 4 dimensions:
  - 3 comprehension strategies (i.e., bridging, elaboration, and paraphrasing)
  - Overall quality
- High imbalance





Models:

- RoBERTa and Multi-Task Learning
- LMMs (Flan-T5 & GPT 3.5)
  - Zero-shot
  - Few-shot
  - Finetuning

Self-Explanation Scoring Cont.

Weighted F1 scores

- Paraphrase: 86.76%
- Elaboration: 89.88%
- Bridging: 79.02%
- Overall: 72.12%

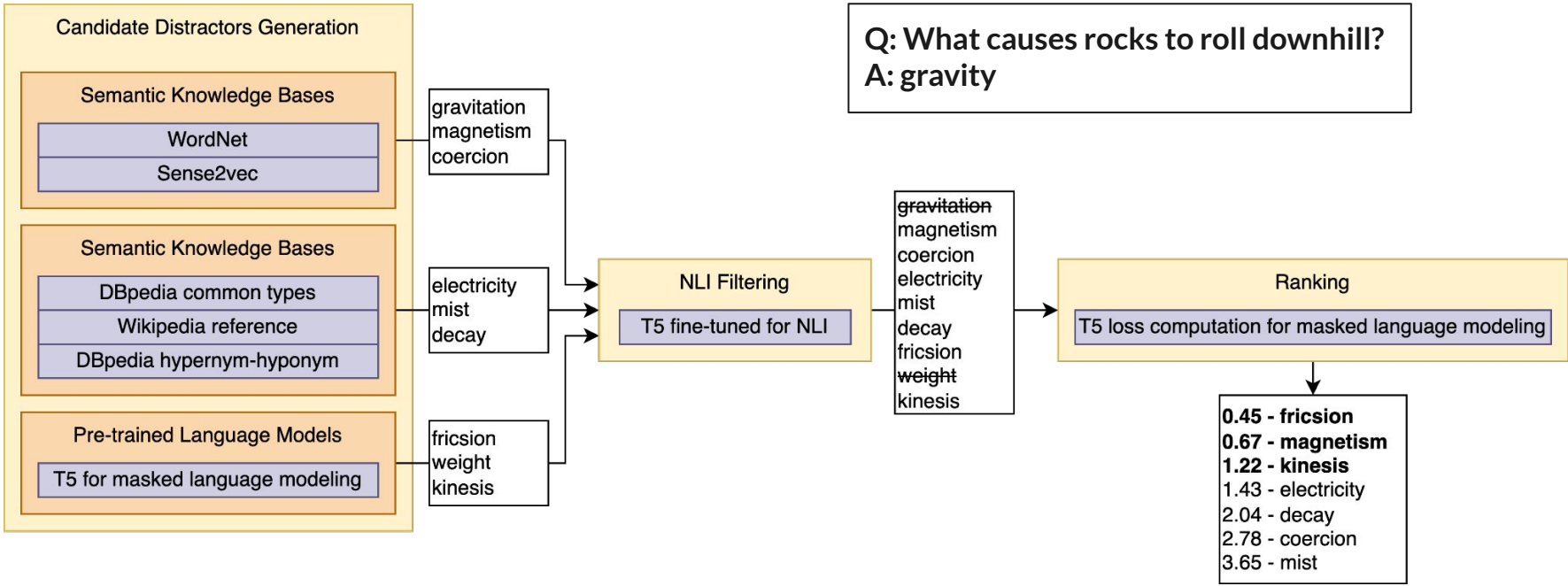
Takeaways

- Good out-of-the-box performance for Paraphrase and Elaboration
  - Really bad (20%-40%) at more complex tasks
- ChatGPT (3.5) had low performance
  - New task, generalization issues
- Fine-tuned Flan-T5 achieved the best results
- What about explainability?

A Shift in the Paradigm - Automated Question Asking

Choose Answers	Generate Questions	Generate distractors
Different approaches: • Named Entities • Supervised based the context • RL based on future QA metrics	• LLM finetuned to generate questions starting from a context and an answer • Further finetuning to control the difficulty of the question	• Generate potential foils with different methods • Filter possible correct answers • Ranking based on QA performance

Better Distractors, Focus on Comprehension



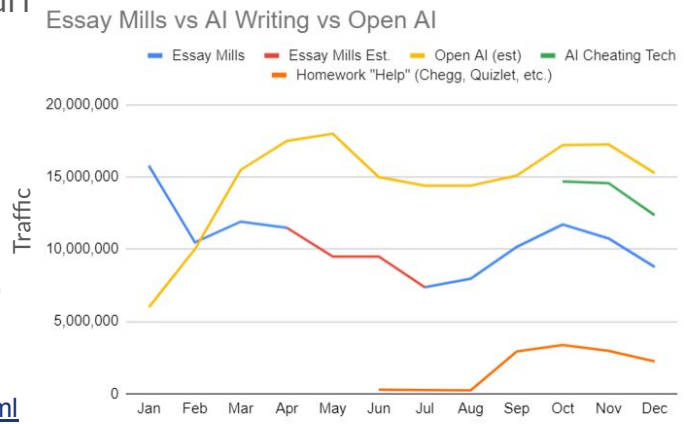
Perspectives for Discussions

Changes in the Educational System

- Adapt to technological advances & prepare for new futures
- Focus on lifelong learning
- Creativity & enhanced critical thinking
- Soft skills & emotional intelligence - human-centered
- Inclusivity & accessibility
- Public-private partnerships

Pitfalls

- Is ChatGPT a Good Teacher Coach?
  - Experiment: Measure zero-shot performance for scoring and providing actionable insights on classroom instruction
  - Outcome: The challenge and importance of generating helpful f
- HITL (Human-In-The-Loop) is a must
  - Hallucinations
- Plagiarism & Cheating
  - No high school cheating increase from ChatGPT [1]
- Will teachers become obsolete?
  - Same question was asked at the ITS Conference 10 years ago



[1] https://www.nytimes.com/2023/12/13/technology/chatbot-cheating-schools-students.html

Skills Needed Today

- Technical proficiency
  - Programming languages (Python, R)
  - Data structures
  - ML/DL algorithms
- Data analytics
- Understanding biases & being aware of ethical implications
- Interdisciplinary knowledge
- Problem solving & critical thinking

Challenges:

- Rapid pace of change
- Skill shortage
- Social implications

Thank you!