

Seminar Natural Language Generation (NLG) — Part 1

Introduction to Natural Language Generation

Henning Wachsmuth

<https://ai.uni-hannover.de>



Motivation

Example: Conversational AI <https://chat.openai.com> (March 31, 2025)

What topics would you cover in a master seminar on natural language generation that has a focus on neural language models? Keep it short.

A master seminar on **Natural Language Generation (NLG) with Neural Language Models** could cover:

1. **Introduction to NLG** – Definition, applications, and challenges
2. **Fundamentals of Neural Language Models** – RNNs, LSTMs, Transformers
3. **Pretraining & Fine-tuning** – GPT, BERT, T5, and instruction tuning
4. **Decoding Strategies** – Greedy search, beam search, nucleus sampling
5. **Evaluation Metrics** – BLEU, ROUGE, METEOR, human evaluation
6. **Bias & Ethical Considerations** – Hallucination, fairness, and safety
7. **Advanced Topics** – Few-shot learning, retrieval-augmented generation (RAG), multimodal NLG
8. **Hands-on Implementation** – Training, fine-tuning, and prompting in practice

Example: Machine translation <https://www.deepl.com> (March 31, 2025)



Translator ▾



Log in

Start free trial

Menu

Translate text
33 languages

Translate files
.pdf, .docx, .pptx

DeepL Write
AI-powered edits

English (detected) ▾

↔ German ▾

Options ▾

A master seminar on Natural Language Generation (NLG) with Neural Language Models could cover: ✕

- Introduction to NLG – Definition, applications, and challenges
- Fundamentals of Neural Language Models – RNNs, LSTMs, Transformers
- Pretraining & Fine-tuning – GPT, BERT, T5, and instruction tuning
- Decoding Strategies – Greedy search, beam search, nucleus sampling
- Evaluation Metrics – BLEU, ROUGE, METEOR, human evaluation
- Bias & Ethical Considerations – Hallucination, fairness, and safety
- Advanced Topics – Few-shot learning, retrieval-augmented generation (RAG), multimodal NLG
- Hands-on Implementation – Training, fine-tuning, and prompting in practice

Ein Masterseminar über die Erzeugung natürlicher Sprache (NLG) mit neuronalen Sprachmodellen könnte folgende Themen behandeln

- Einführung in NLG - Definition, Anwendungen und Herausforderungen
- Grundlagen der neuronalen Sprachmodelle - RNNs, LSTMs, Transformers
- Vortraining und Feinabstimmung - GPT, BERT, T5 und Befehlsabstimmung
- Dekodierungsstrategien - Greedy-Suche, Balkensuche, Kernabstimmung
- Bewertungsmetriken - BLEU, ROUGE, METEOR, menschliche Bewertung
- Voreingenommenheit und ethische Erwägungen - Halluzination, Fairness und Sicherheit
- Fortgeschrittene Themen - Few-shot learning, Retrieval-augmented generation (RAG), multimodale NLG
- Praktische Umsetzung - Training, Feinabstimmung und Prompting in der Praxis

Example: Project Debater https://www.youtube.com/watch?v=UeF_N1r91RQ



Natural Language Generation (NLG)

Natural language processing

■ Computational linguistics (CL)

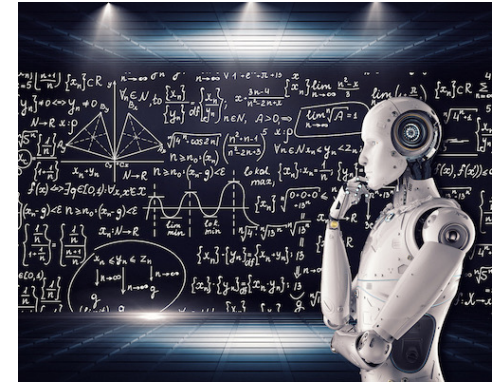
- Intersection of computer science and linguistics
- **Models** for linguistic phenomena, based on knowledge and statistics (machine learning)
- **Technologies** for natural language processing tasks

■ Natural language processing (NLP)

- Methods for understanding and generating speech and human-readable text
- Targets various syntactic, semantic, and pragmatic tasks
- From natural language to structured information, and vice versa

■ Goals of NLP research

- **Creativity.** Novelty of developed models and methods
- **Accuracy.** Effectiveness in tackling tasks
- Empirical research is often seen as stronger than theory

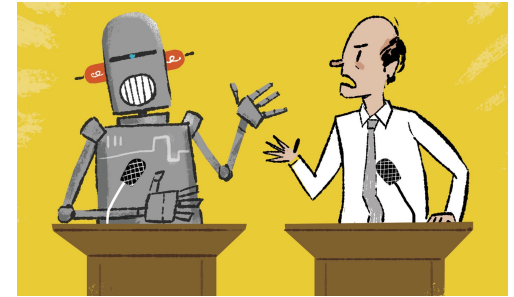


Analysis
Synthesis

Natural language generation

- **Natural language generation (NLG)**

- Methods for the synthesis of natural language (text)
- The goal is to encode structured or semi-structured information in an unstructured text



- **What makes NLG challenging?**

- NLG requires to choose and create a specific textual representation from many potential representations.
- Challenges: Grammaticality, coherence, naturalness, and many more

- **General types of NLG**

- **Data-to-text.** Phrase a new text with data from some knowledge base.
- **Text-to-text.** Rewrite a given text into another text.

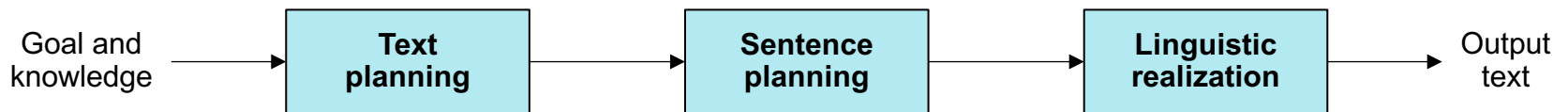
- **General techniques in NLG**

- **Traditional.** Pipeline-like text planning and realization, often knowledge-based
- **End-to-end.** Fine-tuned language modeling, mostly using neural methods

Traditional NLG

▪ **Conceptual NLG process** (Reiter and Dale, 1997)

- **Input.** A goal of what to generate, and knowledge represented in some way
- **Output.** A natural language text



▪ **Main modeling steps**

- **Text planning.** Select content, arrange the discourse structure of sentences
- **Sentence planning.** Aggregate sentence content, make lexical choices, build referring expressions, ...
- **Linguistic realization.** Orthographic, morphological, and syntactic processing

Not all steps are always needed.

Template. "I am
<stance> <issue>,
because <reason>."

Issue. Death penalty
Stance. Con
Reason. "It kills people"

"I am *con* death penalty,
because *the death penalty*
kills people."

Neural language models in NLG

Language model (LM)

- Represents a probability distribution over a sequence of words
- Can be used to predict the most likely next words

Prompt: Can you tell me an argument in favor of having rescue boats?

LM: Without such boats, many innocent refugees will <?>

$P(\text{"die"} \mid \text{dialogue}) = .04$
 $P(\text{"drown"} \mid \text{dialogue}) = .03$
 $P(\text{"suffer"} \mid \text{dialogue}) = .01$

Neural LMs

- Generalize sequences from words to embeddings
- Mostly, probabilities computed based on the neural transformer architecture
- Can be fine-tuned to solve specific tasks or to follow any instruction

”If you wanna hear my view, I think that the EU should allow rescue boats in the Mediterranean Sea. Many innocent refugees will die if there are no such boats. Nothing justifies to endanger the life of innocent people.“

” Without rescue boats, many innocent refugees will die.“

Neural language models: Transformer

Transformer

- A neural encoder-decoder network for parallel processing
- Inputs and outputs in NLP are sequences of (sub-)tokens
- Key concepts: Self-attention and transfer learning

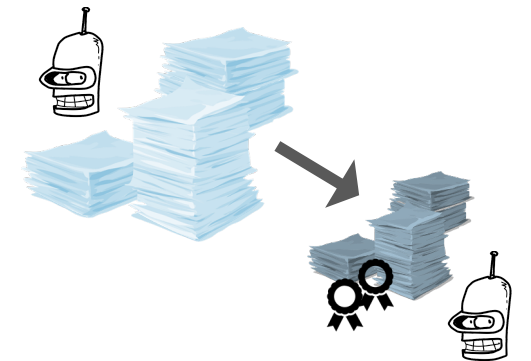
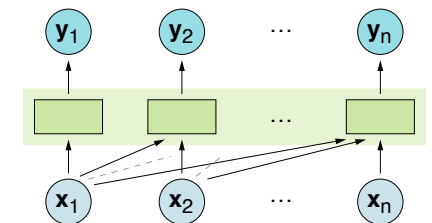
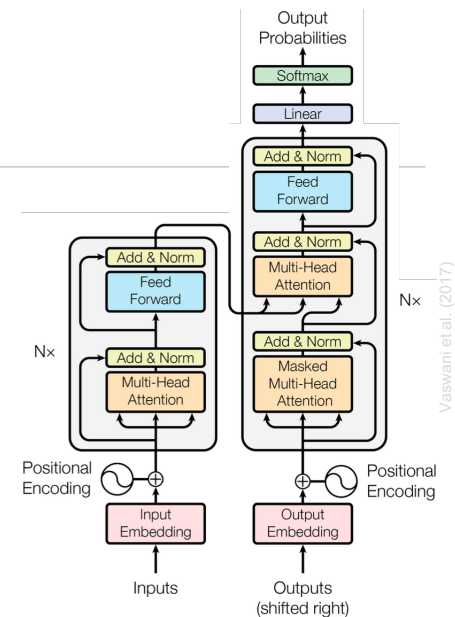
Self-attention

- **Model** each input x_i based on context of other inputs x_j
- Largely solves modeling of long-term input dependencies
- Enables parallelization of input processing

Transfer learning

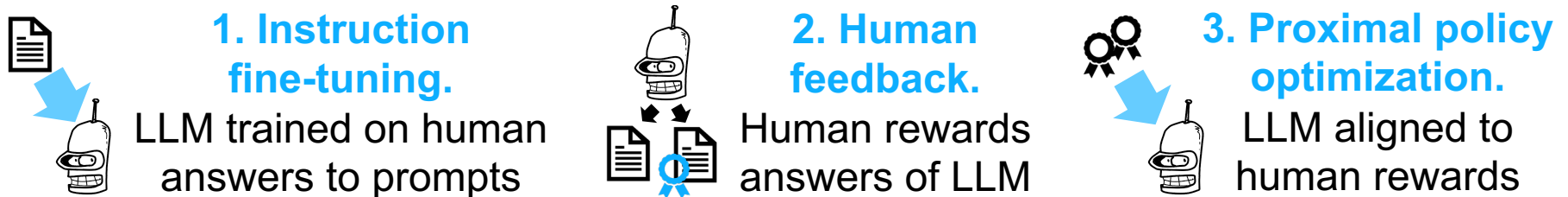
- **Pretrain** network self-supervised on huge text data
- **Fine-tune** it supervised on task-specific training data
- Strongly reduces need for training data

Also, this enables leveraging of knowledge across contexts (more below)



Neural language models: Advanced training methods

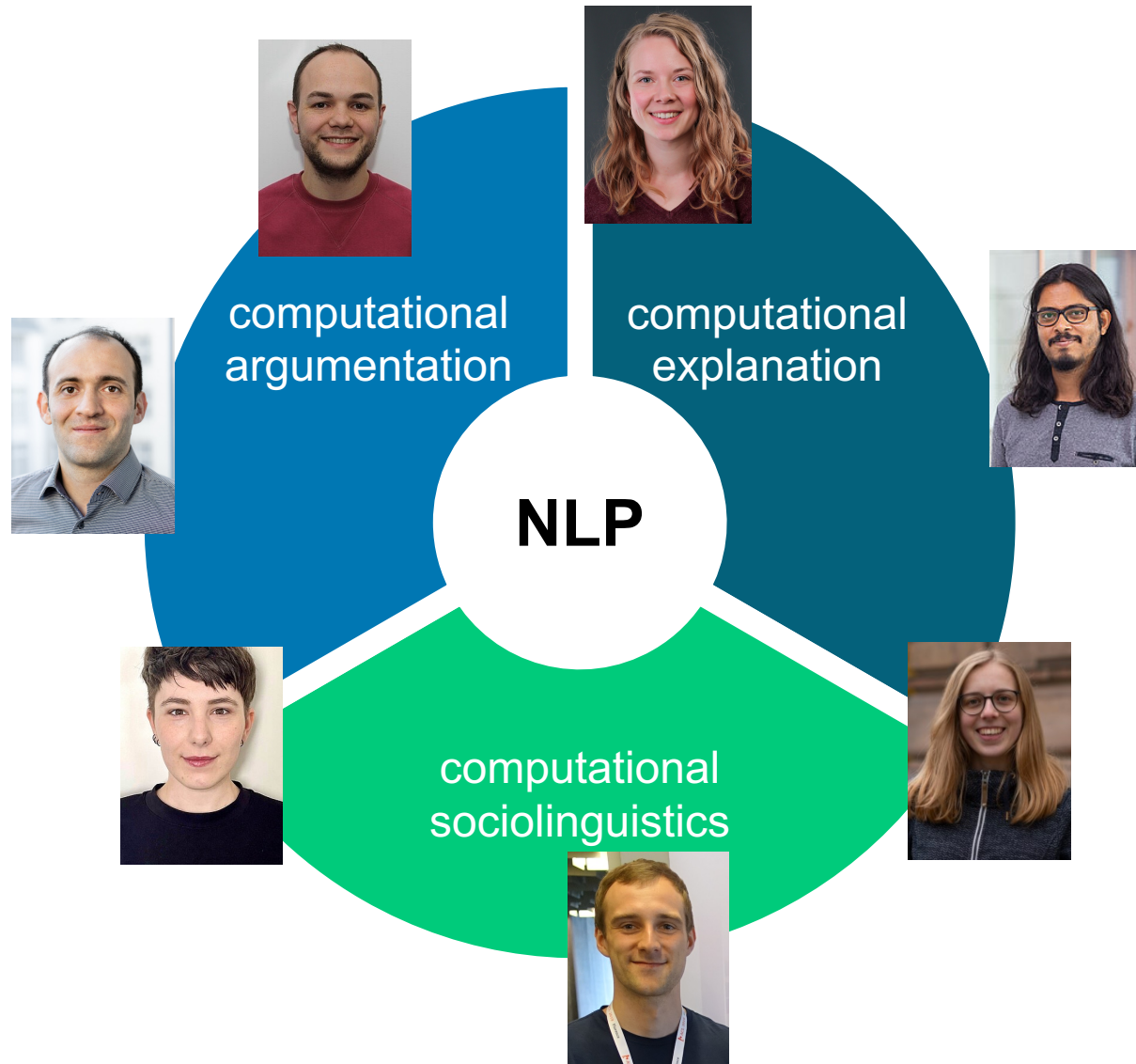
- **Instruction fine-tuning (IFT) and alignment**
 - **IFT.** Train LLMs to create answer-like output to any instruction
 - **Alignment.** Optimize answers towards human-defined preferences
- **Reinforcement learning from human feedback (RLHF)** (Ouyang et al., 2022)
 - Technique for both IFT and alignment



- **Impact**
 - **IFT.** Enables LLMs to tackle any task they are prompted for
 - **Alignment.** Makes answers match certain defined criteria
- **Beyond IFT and alignment**
 - Overparameterization, large input size, in-context learning via prompts

NLG Research of the NLP Group

Natural language generation in the NLP Group



Computational argumentation research

▪ **Argumentation**

- The usage of arguments, along with rhetoric and dialectic, to persuade or agree with others
- Arguments give reasons for claims on controversial issues
- Involved people affect the language and impact of arguments



▪ **Computational argumentation**

- Computational analysis and synthesis of argumentative texts
- Research primarily targets the development of methods for specific tasks
- Important for web search, debating systems, writing support, etc.

▪ **Selected NLG tasks**

- Summarizing argumentative texts
- Synthesizing new claims and arguments
- Improving the quality of arguments

The EU should allow rescue boats...

*”... in the Mediterranean Sea.
Many innocent refugees will die
if there are no rescue boats.“*

Computational sociolinguistics research

▪ Sociolinguistics

- The mutual interactions of society and language
- Relations between social variables and language use
- Language variation across social groups, social contexts, and communicative situations



▪ Computational sociolinguistics

- Analysis and adjustment of datasets and language models in social contexts
- Research primarily targets answers to sociological questions
- Important for human-machine interaction in any sociotechnical system

▪ Selected NLG tasks

- Creating counterfactuals for model training
- Mitigating social and political bias in language
- Generating text that fits a target audience

”Maja dreams of a career as a researcher.“

“Maja works for her career as a researcher.“

Computational explanation research

▪ Explanation

- Explaining is a communicative process between an explainer and an explainee
- Explanations aim at understanding on the explainee's side
- Often has educational or informational purposes



▪ Computational explanation

- Analysis and synthesis of explanatory texts and dialogues
- Research primarily targets the development of methods and systems
- Important for explainability of AI, educational applications, etc.

▪ Selected NLG tasks

- Generating explanations of specific concepts
- Adjusting an explanation to a specific person
- Leading an explanatory dialogue

What's blockchain?

*It's a way that we can trade.
Do you know what trade is?*

*It's when you take turns
doing something, right?*

NLG in this Seminar

This seminar

▪ **Frame of this seminar**

- Basic ideas of natural language generation
- State-of-the-art NLG research
- Connections to research of the NLP Group



▪ **Selected topics covered**

- Argument reconstruction and improvement
- Sociodemographic prompting social bias mitigation
- Generation of explanations and metaphors



▪ **Notice**

- We take a broad view on natural language generation
- Topics are selected according to our research interests
- Basics of NLP are expected rather than taught



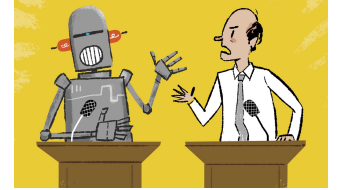
Concrete seminar topics: Next week



Conclusion

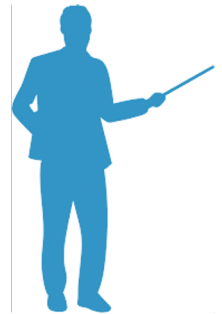
▪ Natural language generation (NLG)

- Computational synthesis of natural language text
- Various methods from templates to neural end-to-end models
- Applications in content creation and human-AI interaction



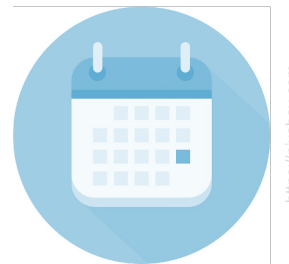
▪ This seminar

- State-of-the-art NLG research around neural language models
- Talks on argumentation, sociolinguistics, and explanation
- Close connection to research in the NLP Group



▪ Next up

- Overview of concrete seminar topics with literature pointers
- Topic preference choice and topic assignment
- Basics of scientific presentation



References

- **Ouyang et al. (2022).** Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Gray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, 2022.
- **Reiter and Dale (1997).** Ehud Reiter and Robert Dale. 1997. Building applied natural language generation systems. *Natural Language Engineering*, 3(1):57–87.
- **Vaswani et al. (2017).** Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser. Attention Is All You Need. In *31st Conference on Neural Information Processing Systems*, 2017.