

Multi-Task Deep Learning in the Legal Domain

Christoph Gebendorfer, Garching, 29.01.2018

Software Engineering betrieblicher Informationssysteme (sebis)
Fakultät für Informatik
Technische Universität München

www.matthes.in.tum.de

- Title: Multi-Task Deep Learning in the Legal Domain
- Author: Christoph Gebendorfer (christoph.gebendorfer@tum.de)
- Supported by: msg systems AG
- Advisor: Ahmed Elnaggar (ahmed.elnaggar@tum.de)
- Start: 15.01.2018
- End: 15.07.2018

Agenda

- Motivation
 - Problem in the Legal Domain
 - Multi-Task Deep Learning
- Research Questions
- Research Approach
- Evaluation Strategy
- Preliminary Results
- Timeline

- Legislative texts
- Regulations
- Enactments
- Patents
- Contracts
- IP documents
- Agreements
- ...



Huge amount of unstructured legal documents and text



Demand for **Natural Language Processing**

Natural Language Processing in the Legal Domain

- Named Entity Recognition
- Named Entity Disambiguation
- Question Answering
- Machine Translation
- Text Classification
- Text Summarization
- Parsing
- Semantic analysis and extraction
- Network analysis, relationship extraction and taxonomy generation

Solutions

- Support Vector Machines
- Random Forests
- Recurrent Neural Networks

However, ...

..., they need two things:

1 Computational Resources

- Task-independent
- Domain-independent



2 Large Annotated Datasets

- Task-dependent
- Domain-dependent



- ▶ Appropriate datasets for tasks in the legal domain are highly limited
... or better phrased, barely exist at all

Unstructured legal
documents and text



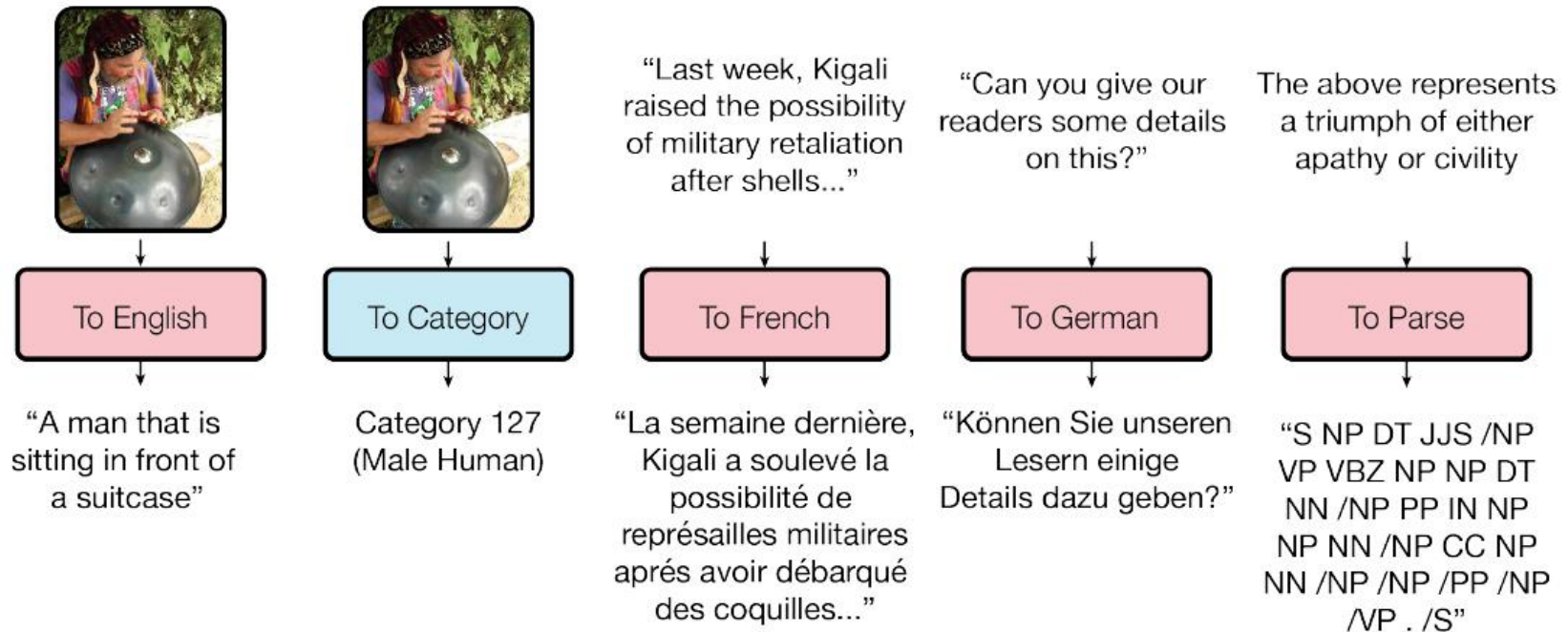
Annotated training
datasets

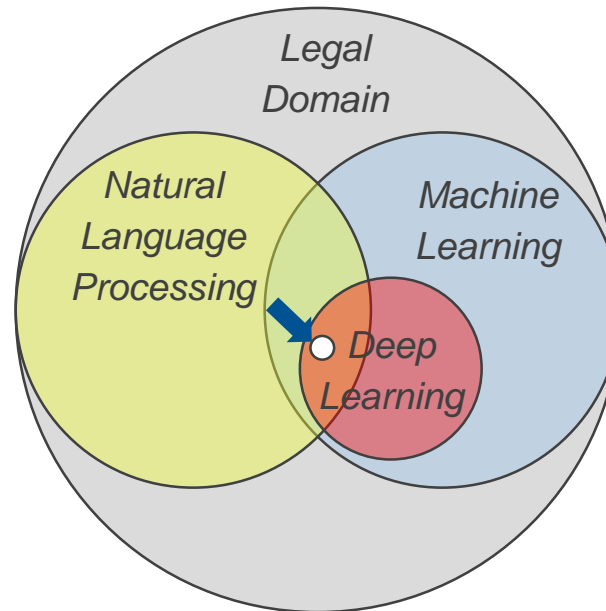
What can we do?

- Creating new datasets
 - Task-dependent
- Use datasets from other domains [1]
 - Pretrain and Adapt
- Active Machine Learning [2] [3]

Are there other methods?

Train **one model** which can perform multiple tasks [4] [5]





Objective:

- Exploit commonalities and overcome task-specific dataset shortage in the legal domain
- Establish Transfer Learning for better results in legal text tasks
- Support generic / task-independent Deep Learning architectures

Agenda

- Motivation
 - Problem in the Legal Domain
 - Multi-Task Deep Learning
- **Research Questions**
- Research Approach
- Evaluation Strategy
- Preliminary Results
- Timeline

1

Can multi-task deep learning be beneficial for tasks in the legal domain?

2

How does training on multiple tasks of the legal domain simultaneously compare to training on each task separately?

3

How far is multi-task deep learning from state-of-the-art solutions in the legal domain?

4

What are good hyperparameters for multi-task deep learning in the legal domain?

5

Can feeding datasets from other domains improve the performance of tasks in the legal domain?

Agenda

- Motivation
 - Problem in the Legal Domain
 - Multi-Task Deep Learning
- Research Questions
- **Research Approach**
- Evaluation Strategy
- Preliminary Results
- Timeline



Deductive Reasoning

- Search for datasets in the legal domain ✓
- Choose a Multi-Task model ✓
- Integrate datasets into the Multi-Task model ⚙️
- Conduct experiments
 - Train the Multi-Task model on legal tasks —
 - Generate information according to the hypotheses —
- Evaluate generated information —



Backed by literature research



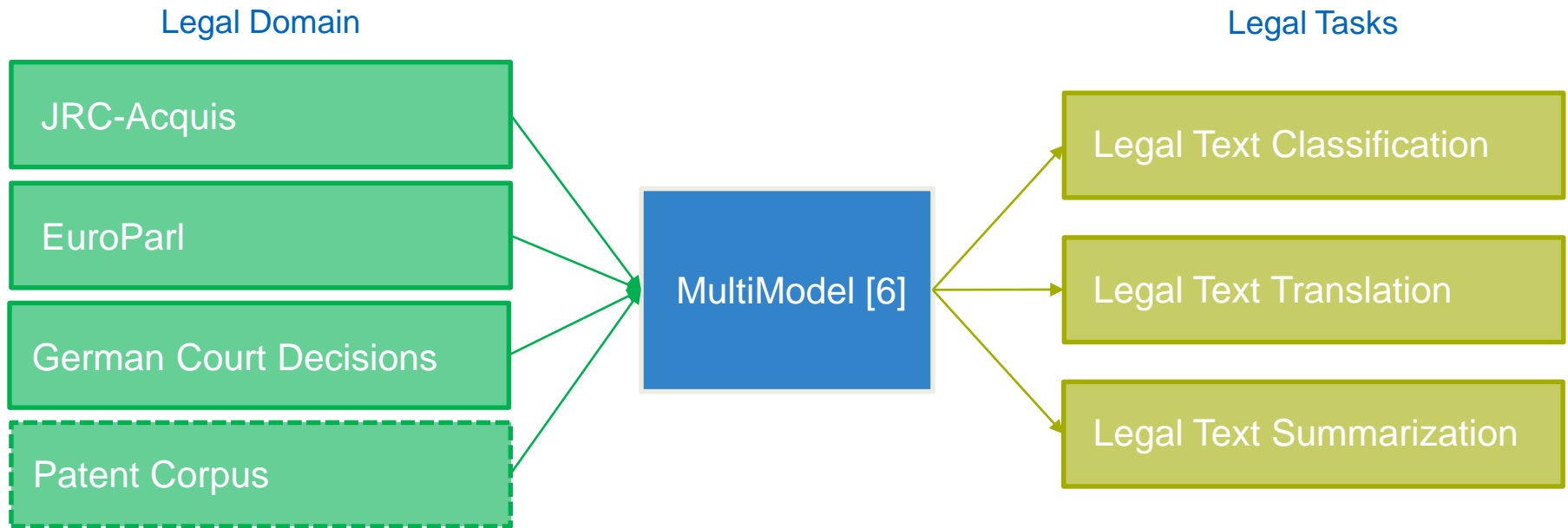
Verify or disprove research questions

Research Approach - Corpora

Corpus	LD	EN	DE	Classification	Summarization	Parsing	Cost	Size
JRC-Acquis	X	X	X	X	X	-	-	+++
The HOLJ Corpus	X	X	-	-	X	-	-	+
Patent Decisions	X	X	-	X	-	-	-	?
Reuters Newswire	~	X	X	X	~	-	-	++
CNN / Daily Mail	~	X	-	X	X	-	-	+++
Annotated Gigaword	~	X	-	-	X	X	X	++++
DeReKo	~	-	X	-	-	-	-	++++
DUC Corpora	~	X	-	-	X	-	-	+
Huge German Corpus	~	-	X	-	-	X	-	+++
NEGR@ Corpus	~	-	X	-	-	X	-	++
TIGER corpus	~	-	X	-	-	X	-	+++

Research Approach - Corpora

Corpus	LD	EN	DE	Classification	Summarization	Parsing	Cost	Size
JRC-Acquis	X	X	X	X	X	-	-	+++
EuroParl	X	X	X	-	-	-	-	+++
German Court Decisions	X	-	X	X	X	-	-	++
Patent Decisions	X	X	-	X	-	-	-	?
Reuters Newswire	~	X	X	X	~	-	-	++
CNN / Daily Mail	~	X	-	X	X	-	-	+++
Annotated Gigaword	~	X	-	-	X	X	X	++++
DeReKo	~	-	X	-	-	-	-	++++
DUC Corpora	~	X	-	-	X	-	-	+
Huge German Corpus	~	-	X	-	-	X	-	+++
NEGR@ Corpus	~	-	X	-	-	X	-	++
TIGER corpus	~	-	X	-	-	X	-	+++



	Translation	Classification	Summarization
JRC-Acquis	X	X	X
EuroParl	X		
German Court Decisions		X	X
Patent Corpus		X	

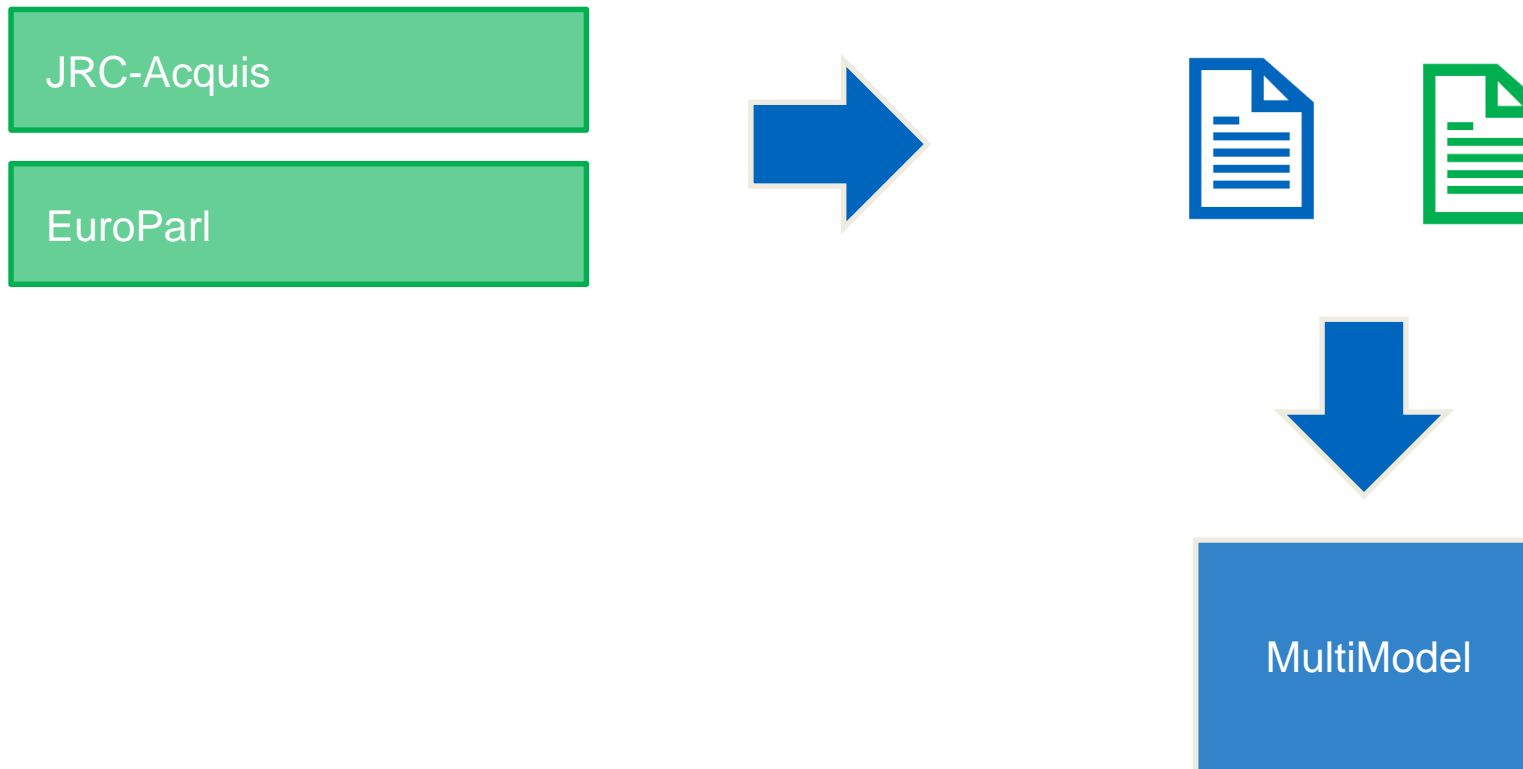
- Motivation
 - Problem in the Legal Domain
 - Multi-Task Deep Learning
- Research Questions
- Research Approach
- **Evaluation Strategy**
- Preliminary Results
- Timeline

Tasks	Text Classification	Text Summarization	Machine Translation
Measure	F1 Score $2 * \frac{\textit{precision} * \textit{recall}}{\textit{precision} + \textit{recall}}$	ROGUE-1 ROGUE-2 <i>Overlap of n grams between system and reference summaries</i>	BLEU Score <i>Overlap of n grams between system and reference translation according to total appearance</i>

- Motivation
 - Problem in the Legal Domain
 - Multi-Task Deep Learning
- Research Questions
- Research Approach
- Evaluation Strategy
- **Preliminary Results**
- Timeline

- XML format
- Metainformation
- Tagged paragraphs

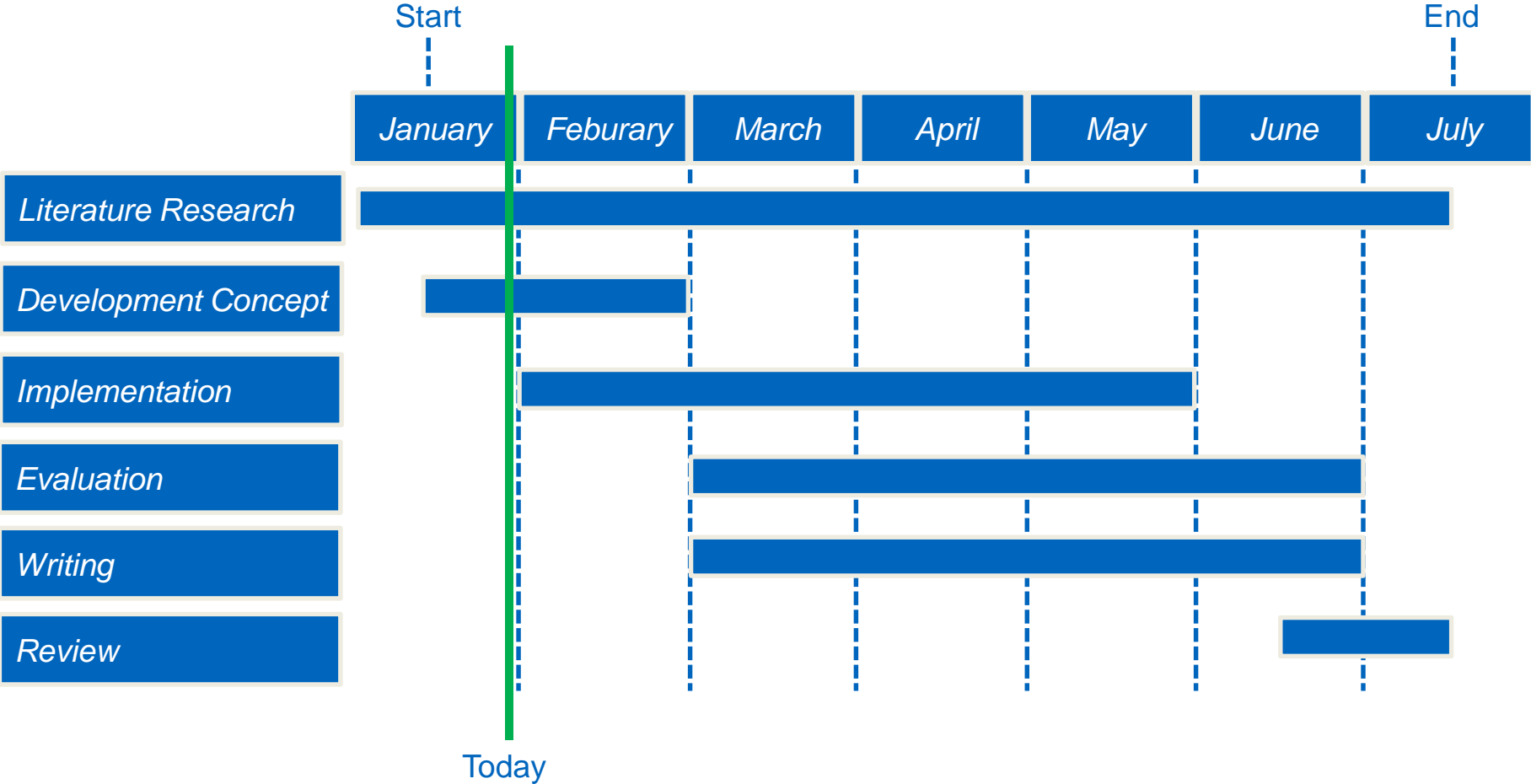
- MOSES format
- Aligned text line by line
- One file per language
- cs, de, en, es, fr, it, sv



Agenda

- Motivation
 - Problem in the Legal Domain
 - Multi-Task Deep Learning
- Research Questions
- Research Approach
- Evaluation Strategy
- Preliminary Results
- **Timeline**

Timeline



- [1] Robin Otto. Transfer Learning for Name Entity Linking with Deep Learning. Technical University Munich, 2018
- [2] Bernhard Walzl, Johannes Muhr, Ingo Glaser, Georg Bonczek, Elena Scepankova and Florian Matthes. Classifying Legal Norms with Active Machine Learning. JURIX, Luxembourg, 2017
- [3] Micheal Legenc. Using Natural Language Processing and Machine Learning to Assist First-Level Customer Support for Contract Management. Technical University Munich, 2018
- [4] Rich Caruana. Multitask Learning. Machine Learning, 28(1):41-75, 1997
- [5] Iman Jundi. Using Multitask Deep Learning for Question Answering: A Use Case on Insurance Question Answering Dataset. Technical University Munich, 2018
- [6] Lukasz Kaiser, Aidan N. Gomez, Noam Shazeer, Ashish Vaswani, Niki Parmar, Llion Jones and Jakob Uszkoreit. One Model To Learn Them All. CoRR, abs/1706.05127, 2017. URL <http://arxiv.org/abs/1706.05137>

1

What preprocessing of the datasets is necessary for translation, classification and summarization?

2

What factors should be considered when training Deep Learning / Multi-Task models?

3

What impact do dataset properties have on the performance?

4

What state-of-the-art solutions exist for which tasks? Are they domain-dependent?

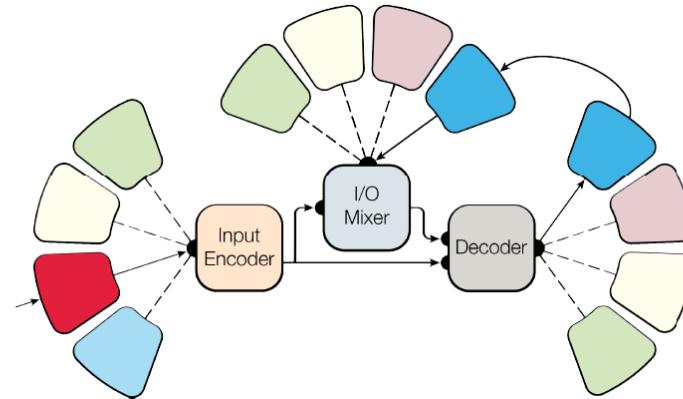
5

What solutions are currently used in practice? Are they competitive?

6

How can the MultiModel be extended?

Architecture



Building Blocks

