Neuro-Symbolic Modeling: Looking back at the semester

THE UNIVERSITY OF UTAH

Neural networks the default modeling tool today

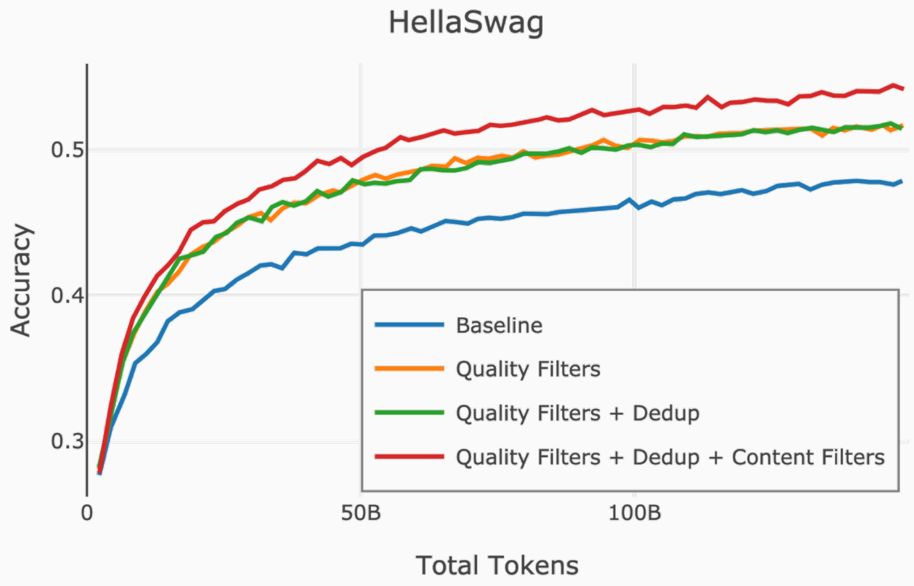
They have demonstrated remarkable successes

Neural networks the default modeling tool today

They have demonstrated remarkable successes

But…they need a lot of data!

Source	Doc Type	UTF-8 bytes *(GB)*	Documents *(millions)*	Unicode words *(billions)*	Llama tokens *(billions)*
Common Crawl	web pages	9,812	3,734	1,928	2,479
GitHub	code	1,043	210	260	411
Reddit	social media	339	377	72	89
Semantic Scholar	papers	268	38.8	50	70
Project Gutenberg	books	20.4	0.056	4.0	6.0
Wikipedia, Wikibooks	encyclopedic	16.2	6.2	3.7	4.3
Total		**11,519**	**4,367**	**2,318**	**3,059**



Luca Soldaini et al. "Dolma: An Open Corpus of Three Trillion Tokens for Language Model Pretraining Research." ACL 2024

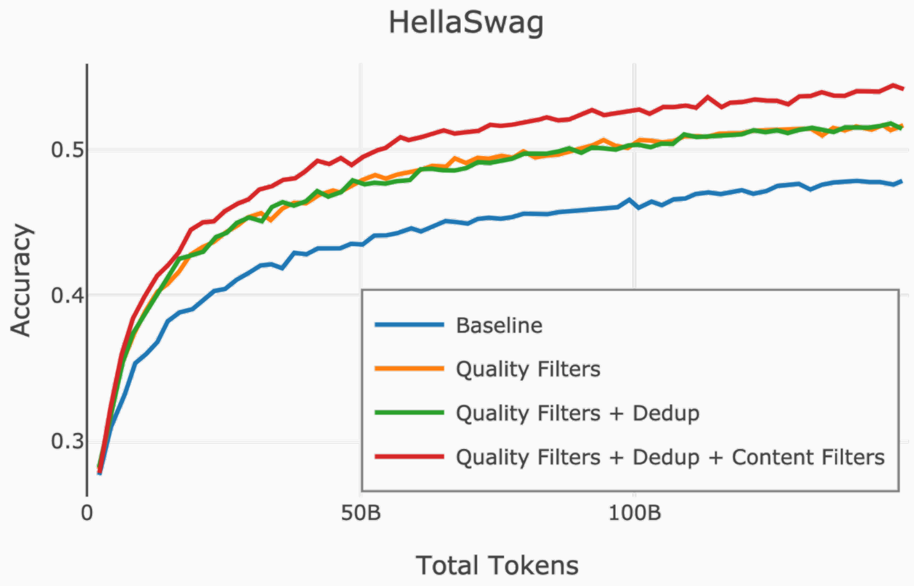
Neural networks the default modeling tool today

They have demonstrated remarkable successes

But…they need a lot of data!

Data can be a precious resource

Source	Doc Type	UTF-8 bytes (GB)	Documents (millions)	Unicode words (billions)	Llama tokens (billions)
Common Crawl	web pages	9,812	3,734	1,928	2,479
GitHub	code	1,043	210	260	411
Reddit	social media	339	377	72	89
Semantic Scholar	papers	268	38.8	50	70
Project Gutenberg	books	20.4	0.056	4.0	6.0
Wikipedia, Wikibooks	encyclopedic	16.2	6.2	3.7	4.3
Total		**11,519**	**4,367**	**2,318**	**3,059**



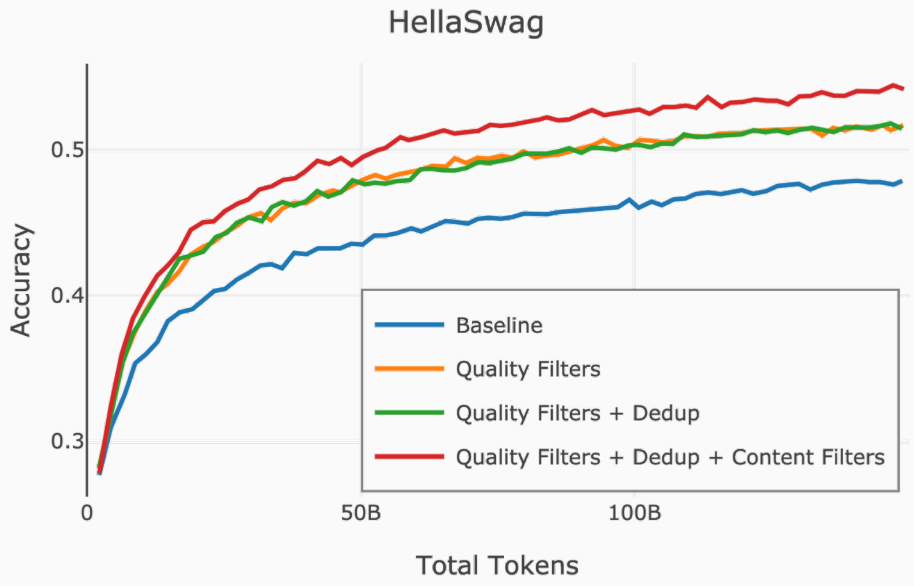
Luca Soldaini et al. "Dolma: An Open Corpus of Three Trillion Tokens for Language Model Pretraining Research." ACL 2024

Neural networks the default modeling tool today

They have demonstrated remarkable successes

But…they need a lot of data!

Source	Doc Type	UTF-8 bytes *(GB)*	Documents *(millions)*	Unicode words *(billions)*	Llama tokens *(billions)*
Common Crawl	web pages	9,812	3,734	1,928	2,479
GitHub	code	1,043	210	260	411
Reddit	social media	339	377	72	89
Semantic Scholar	papers	268	38.8	50	70
Project Gutenberg	books	20.4	0.056	4.0	6.0
Wikipedia, Wikibooks	encyclopedic	16.2	6.2	3.7	4.3
Total		**11,519**	**4,367**	**2,318**	**3,059**



Data can be a precious resource

Let's see an example

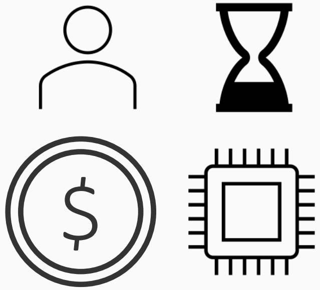
Luca Soldaini et al. "Dolma: An Open Corpus of Three Trillion Tokens for Language Model Pretraining Research." ACL 2024

Whom does NLP serve?

Massive models and data

Improvements in benchmark performance…

…benefiting everyone?



Low resource situations?

Low or no data problems
Long tail languages
Limited compute resources
Limited number of users

Example: AI for mental health

Low data + legal usage limitations

What are low resource situations?

The usual answer: limited data, limited compute

An alternative perspective: What leads to data and compute limitations?

Whose data and compute is typically "low resource"?

- Historically underserved regions of the world
- Historically underserved people everywhere
- Languages and domains that are not financially "interesting"
- Groups that often have legal protections

Not a mutually exclusive categorization

The problem with purely data-driven machine learning

Large models trained on large amounts of data can functionally approximate the intelligent behavior that led to the creation of the data

- Eg: Language models can produce token sequences that resemble language in style and content

We need AI systems that…

- …can be controlled via mechanisms other than just “add more data”
- …can provide insight into their reasoning processes

Are we modeling problems in their full richness?

Or: are we modeling for benchmark test sets?

Challenge: Modeling strategies that depend on or expose a theory to reduce data dependence

(or at least assume the existence of a theory)

- Otherwise, we are just guessing equations involving high dimensional spaces, without understanding what they mean

- Thales, versus a modern understanding of eclipses

  Precise predictions via curve fitting versus a philosophy of the underlying phenomenon

Symbolic Artificial Intelligence

An agenda for artificial intelligence that focuses on knowledge representation and reasoning

Uses symbols (i.e. *discrete* labels), rules, and logic to represent knowledge

- Typically hand crafted rules

Well understood algorithms can reason about it

Rich human-auditable representations of the semantics and behavior of the programs

The problem with GOFAI

Worked well, but *only* for a very small set of examples

Did not handle uncertainty and noise well

Took several years of human effort to create

And still did not generalize

Because too many hand-crafted rules

Hand-crafted rules do not really reflect how complex phenomena like language and vision work in practice!

Case study: AI for mental health

Trigger warning: Discussion about and lightly edited example of self-harm ahead

How is this a low resource situation?

Plenty of interest in AI for mental health

- Even more so in AI for health

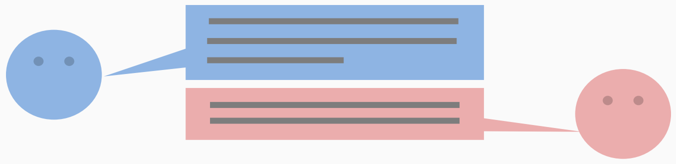
But...

- Mental health issues are often a taboo topic worldwide, and found data (e.g. by web scraping) does not reflect the struggles of patients

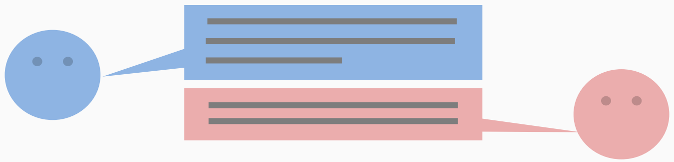
- Unclear whether standard datasets represent neurodivergence well enough

- Actual counseling data can be behind legal protections, prohibiting API-based LLM access and cloud GPU infrastructure

A trip to a therapist is…

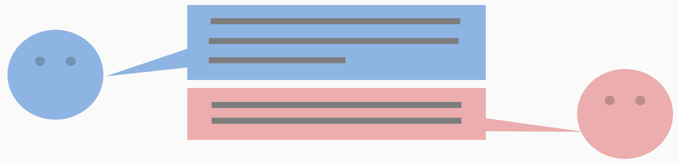


A trip to a therapist is...

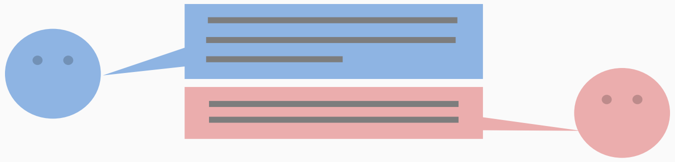


A trip to a therapist is…

Diagnosis → Treatment → Post-treatment evaluation



A trip to a therapist is…



Essentially a dialogue between two people

- Can we guide this dialogue with automated helpers?
- Are some therapists or treatments better than others?
- Can we measure patient improvements?

SAFE UT Crisis counseling is a whole different game

Users can anonymously chat with a crisis counselor

- Like a suicide hotline, but via text

Original and primary audience: high school students

- Recently expanded to frontline workers, national guard recently

https://safeut.org/

SAFE UT Crisis counseling is a whole different game

Users can anonymously chat with a crisis counselor

- Like a suicide hotline, but via text

Original and primary audience: high school students

- Recently expanded to frontline workers, national guard recently

Easy to roll out to users, but not so easy to hire trained counselors

https://safeut.org/

SAFE UT Crisis counseling is a whole different game

Users can anonymously chat with a crisis counselor

- Like a suicide hotline, but via text

Original and primary audience: high school students

- Recently expanded to frontline workers, national guard recently

Easy to roll out to users, but not so easy to hire trained counselors

Text-based counseling (crisis or otherwise) breaks new ground in psychotherapy

- A single counselor sometimes has seven simultaneous conversations with at-risk people!
- Same user may be chatting with different counselors over a day or two

https://safeut.org/

SAFE UT Crisis counseling is a whole different game

Users can anonymously chat with a crisis counselor
- Like a suicide hotline, but via text

Original and primary audience: high school students
- Recently expanded to frontline workers, national guard recently

Easy to roll out to users, but not so easy to hire trained counselors

Text-based counseling (crisis or otherwise) breaks new ground in psychotherapy
- A single counselor sometimes has seven simultaneous conversations with at-risk people!
- Same user may be chatting with different counselors over a day or two

Can NLP help?

https://safeut.org/

An anonymized high risk session

The whole session represents a high risk patient…

…but it is individual utterances that reveal useful cues for a counselor

I’ve been thinking of ending my life, my mental health is bad. I had an attempt [#DATE].

I am here for you. How can I help?

Hi. I feel I’m worthless and unhappy all the time. And I’m tired of being anxious.

I am sorry you feel that way. Have you talked to a therapist about your anxiety?

I have. But I can’t work with my therapist anymore. I’m looking for a new one.

Have you ever tried things like the 5 senses grounding skill for your anxiety?

I did. I tried so hard, and nothing worked. I’m just exhausted. Sometimes I feel like wanting to give it an end.

⋮

Mattia Medina Grespan, Meghan Broadbent, Xinyao Zhang, Katherine Axford, Brent Kious, Zac Imel, and Vivek Srikumar. Logic-driven Indirect Supervision: An Application to Crisis Counseling. In ACL 2023.

An anonymized high risk session

The whole session represents a high risk patient…

…but it is individual utterances that reveal useful cues for a counselor

🧑	I've been thinking of ending my life, my mental health is bad. I had an attempt [#DATE].	Lifetime ideation Prior attempts
👩‍⚕️	I am here for you. How can I help?	
🧑	Hi. I feel I'm worthless and unhappy all the time. And I'm tired of being anxious.	No relevant information for the crisis counselor
👩‍⚕️	I am sorry you feel that way. Have you talked to a therapist about your anxiety?	
🧑	I have. But I can't work with my therapist anymore. I'm looking for a new one.	No relevant information for the crisis counselor
👩‍⚕️	Have you ever tried things like the 5 senses grounding skill for your anxiety?	
🧑	I did. I tried so hard, and nothing worked. I'm just exhausted. Sometimes I feel like wanting to give it an end.	Passive ideation

⋮

Mattia Medina Grespan, Meghan Broadbent, Xinyao Zhang, Katherine Axford, Brent Kious, Zac Imel, and Vivek Srikumar. Logic-driven Indirect Supervision: An Application to Crisis Counseling. In ACL 2023.

An anonymized high risk session

The whole session represents a high risk patient…

…but it is individual utterances that reveal useful cues for a counselor

🧑	I've been thinking of ending my life, my mental health is bad. I had an attempt [#DATE].	D_1, D_9	Lifetime ideation Prior attempts
👩‍⚕️	I am here for you. How can I help?		
🧑	Hi. I feel I'm worthless and unhappy all the time. And I'm tired of being anxious.	No-D	No relevant information for the crisis counselor
👩‍⚕️	I am sorry you feel that way. Have you talked to a therapist about your anxiety?		
🧑	I have. But I can't work with my therapist anymore. I'm looking for a new one.	No-D	No relevant information for the crisis counselor
👩‍⚕️	Have you ever tried things like the 5 senses grounding skill for your anxiety?		
🧑	I did. I tried so hard, and nothing worked. I'm just exhausted. Sometimes I feel like wanting to give it an end.	D_{2b}	Passive ideation

⋮

Mattia Medina Grespan, Meghan Broadbent, Xinyao Zhang, Katherine Axford, Brent Kious, Zac Imel, and Vivek Srikumar. Logic-driven Indirect Supervision: An Application to Crisis Counseling. In ACL 2023.

An anonymized high risk session

The whole session represents a high risk patient…

…but it is individual utterances that reveal useful cues for a counselor

I've been thinking of ending my life, my mental health is bad. I had an attempt [#DATE].	D_1, D_9	Lifetime ideation Prior attempts
I am here for you. How can I help?		
Hi. I feel I'm worthless and unhappy all the time. And I'm tired of being anxious.	No-D	No relevant information for the crisis counselor
		levant information for the counselor
an end.	D_{26}	ive ideation

⋮

Such utterance level labels have been shown to be useful to train counselors by zeroing in on nuances during the conversation.

Can we build utterance level models?

Mattia Medina Grespan, Meghan Broadbent, Xinyao Zhang, Katherine Axford, Brent Kious, Zac Imel, and Vivek Srikumar. Logic-driven Indirect Supervision: An Application to Crisis Counseling. In ACL 2023.

I've been thinking of ending my life, my mental health is bad. I had an attempt [#DATE].

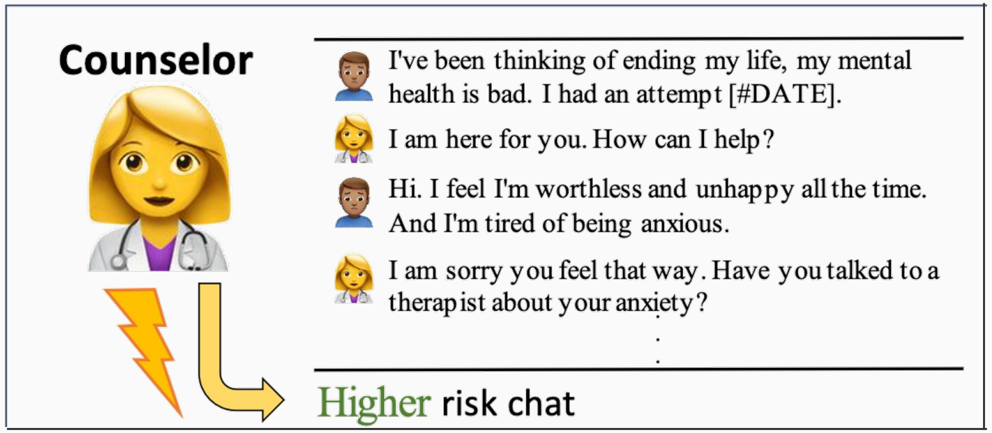
I am here for you. How can I help?

Hi. I feel I'm worthless and unhappy all the time. And I'm tired of being anxious.

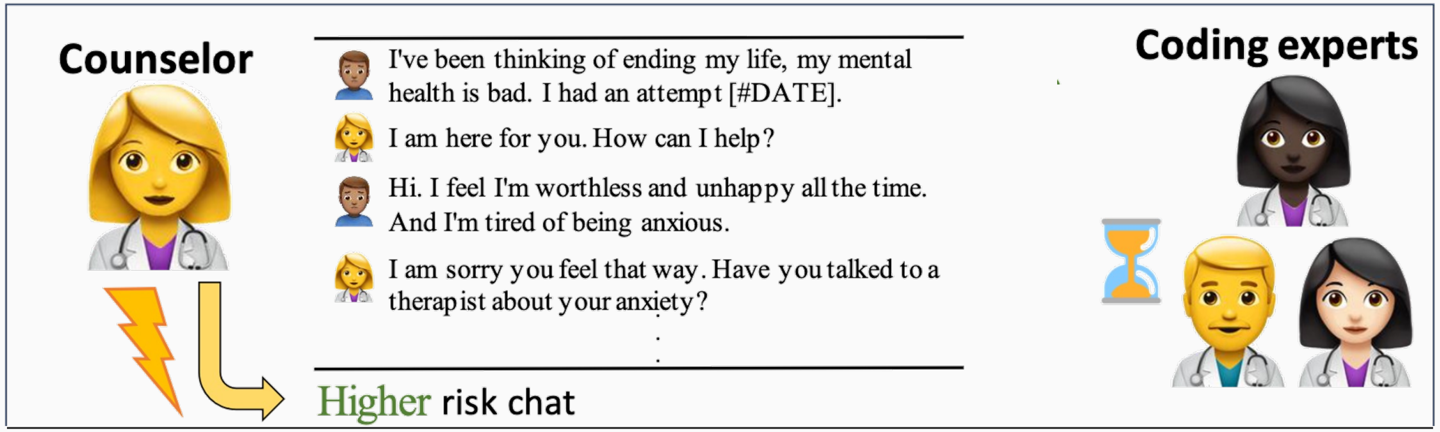
I am sorry you feel that way. Have you talked to a therapist about your anxiety?

⋮

Easy to obtain session-level labeled data: A counselor can quickly tag a session as part of their workflow

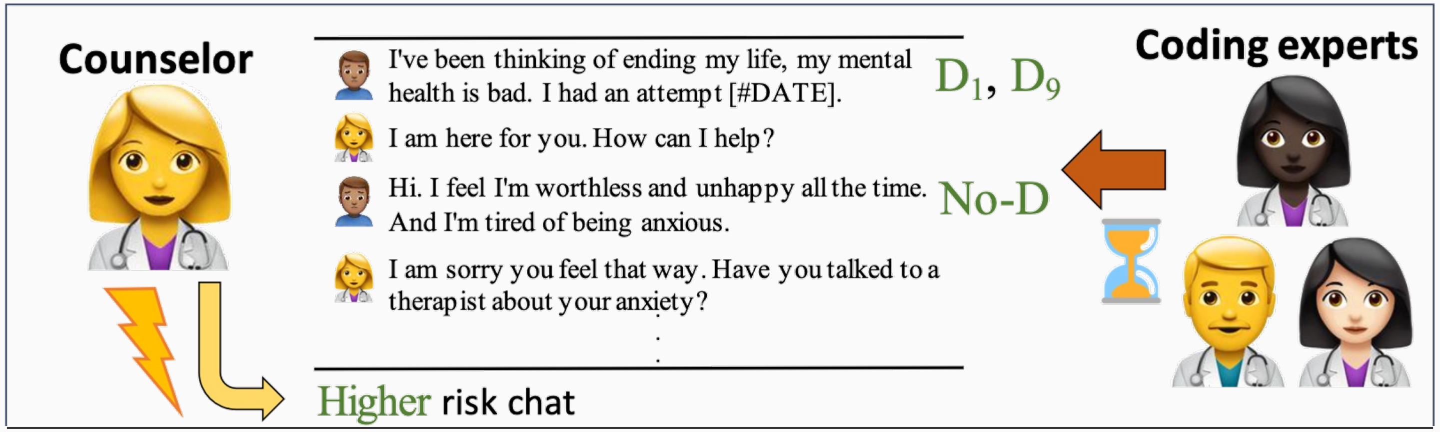


Easy to obtain session-level labeled data: A counselor can quickly tag a session as part of their workflow



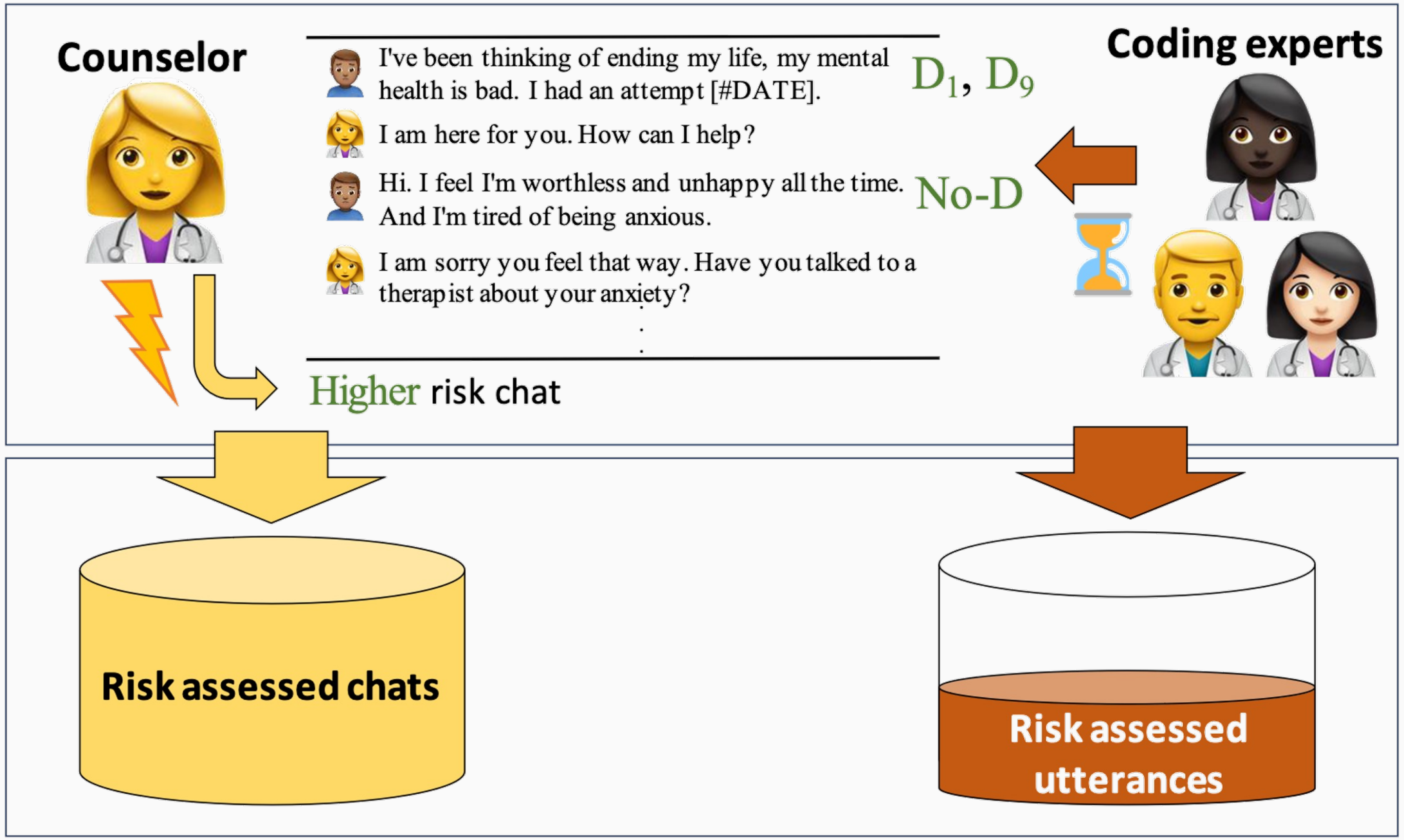
Utterance-level data is much harder to obtain: Needs post hoc expert annotation

Easy to obtain session-level labeled data: A counselor can quickly tag a session as part of their workflow



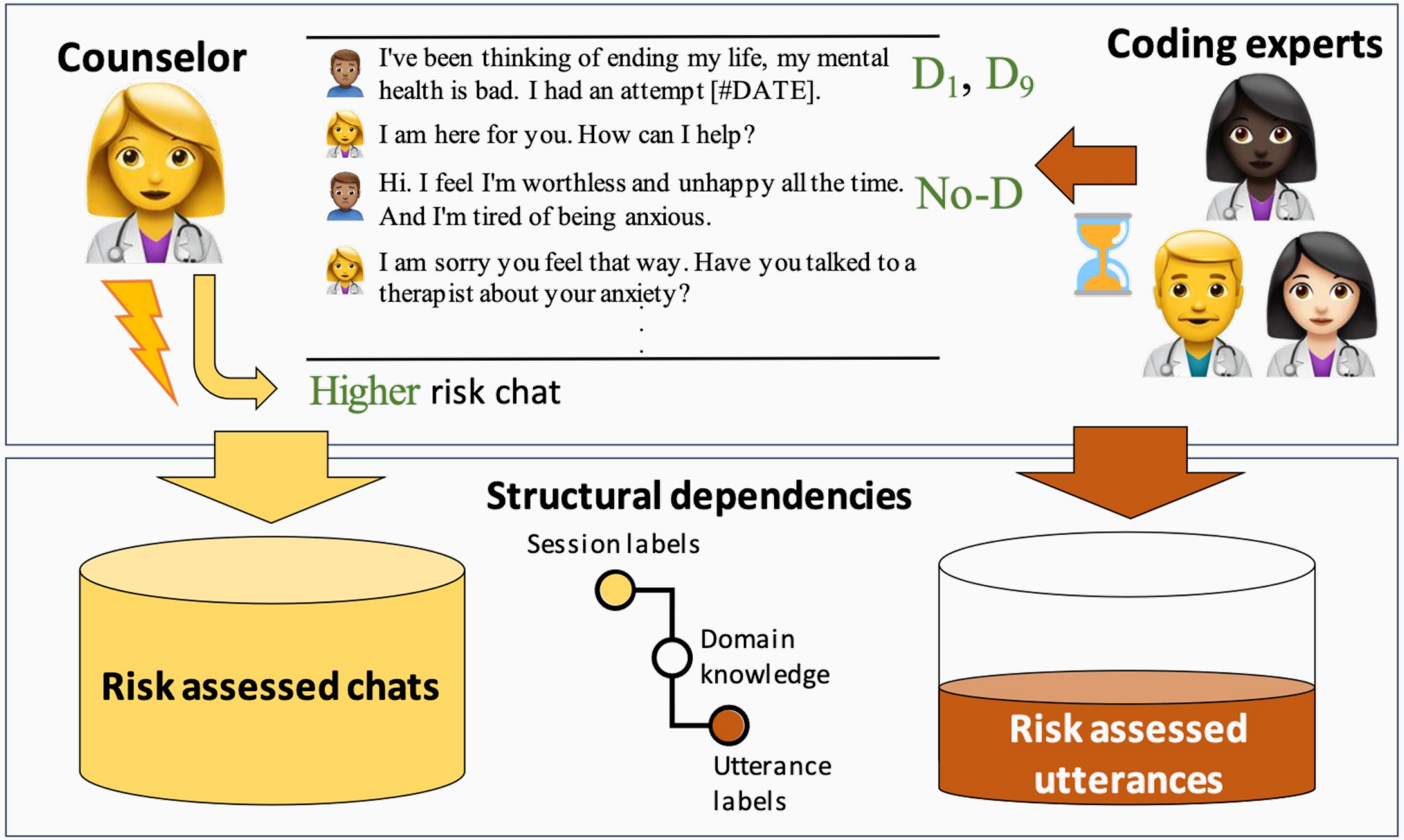
Utterance-level data is much harder to obtain: Needs post hoc expert annotation

Easy to obtain session-level labeled data: A counselor can quickly tag a session as part of their workflow



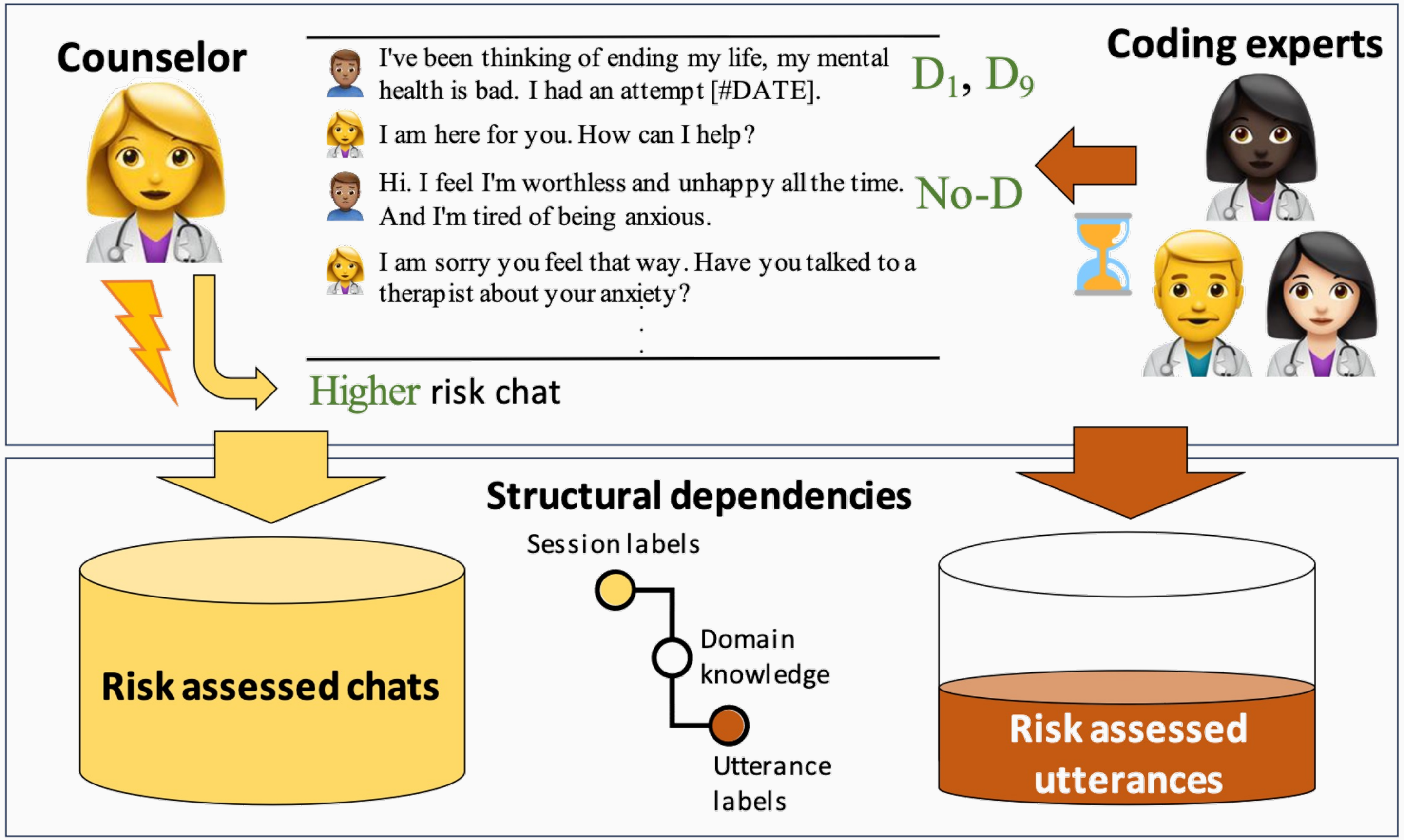
Utterance-level data is much harder to obtain: Needs post hoc expert annotation

Easy to obtain session-level labeled data: A counselor can quickly tag a session as part of their workflow

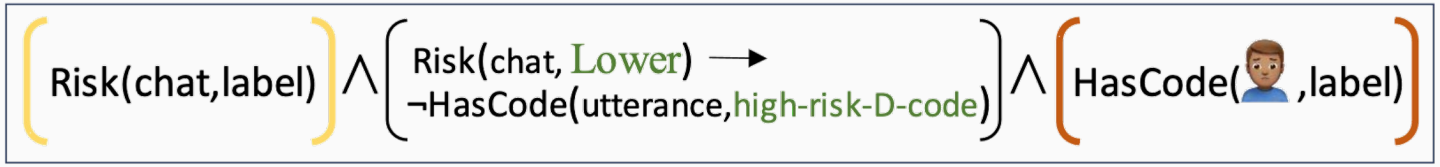


Utterance-level data is much harder to obtain: Needs post hoc expert annotation

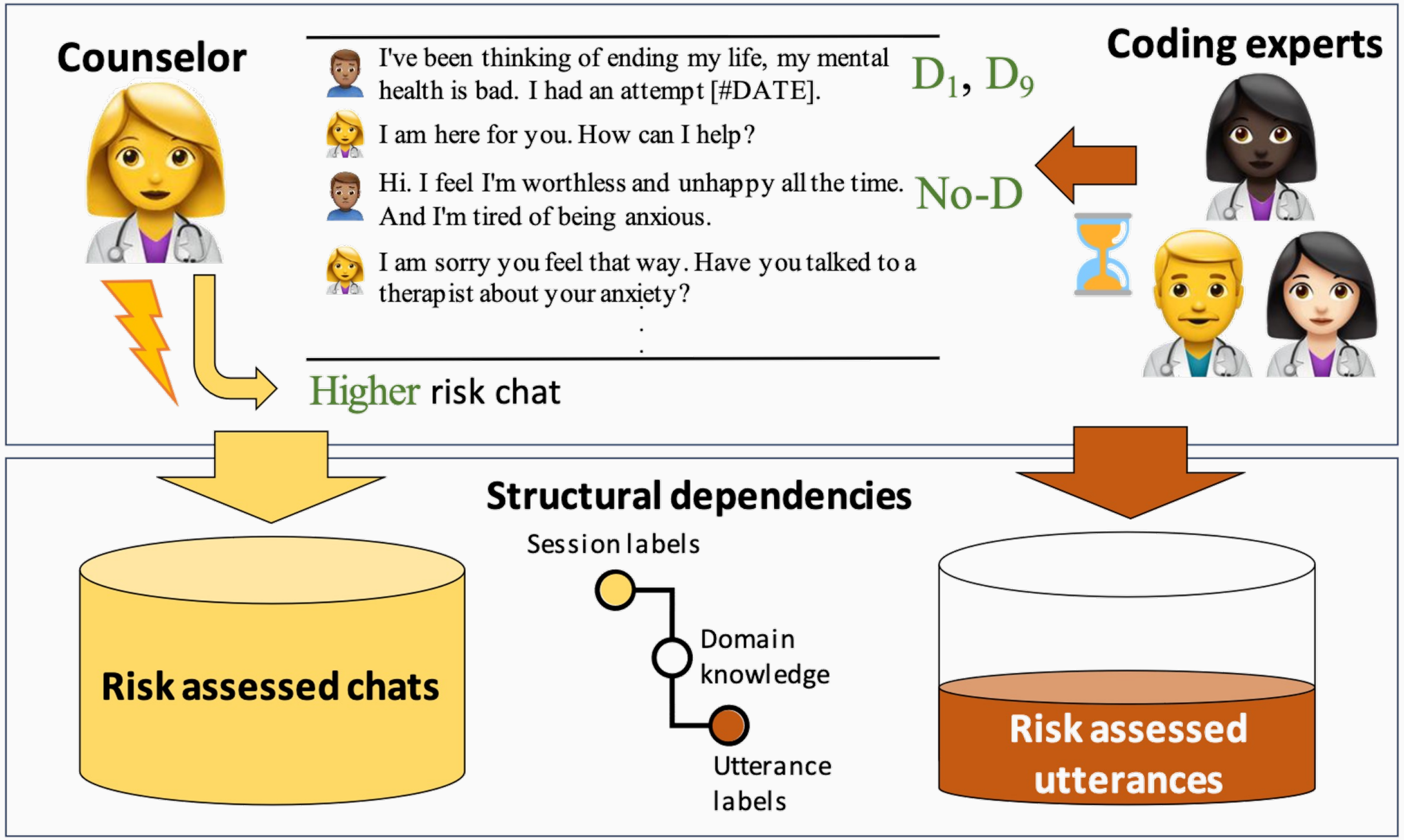
Easy to obtain session-level labeled data: A counselor can quickly tag a session as part of their workflow



Utterance-level data is much harder to obtain: Needs post hoc expert annotation



Easy to obtain session-level labeled data: A counselor can quickly tag a session as part of their workflow

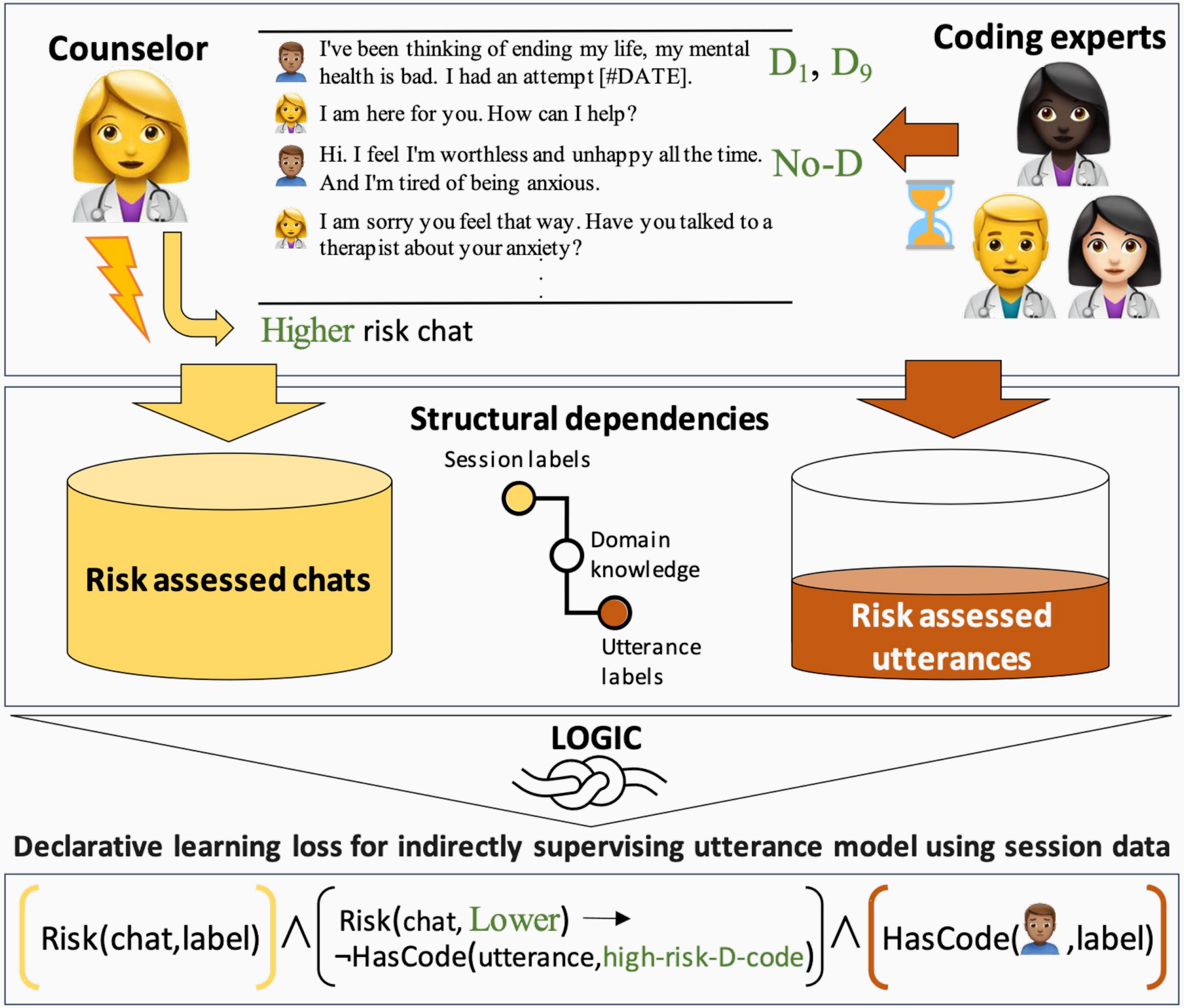


Utterance-level data is much harder to obtain: Needs post hoc expert annotation

Can the larger amounts of session-level data help supervise an utterance-level model?

Risk(chat,label) ∧ (Risk(chat, Lower) ⟶ ¬HasCode(utterance,high-risk-D-code)) ∧ (HasCode(👨🏽,label))

Easy to obtain session-level labeled data: A counselor can quickly tag a session as part of their workflow



Utterance-level data is much harder to obtain: Needs post hoc expert annotation

Key idea: We can compile such declarative rules into losses that we can then use with any model and optimizer

Neural networks versus Symbolic AI

Neural networks

The best pattern recognition engines today

Symbolic AI

Historically well developed and has deep algorithmic understanding

Neural networks versus Symbolic AI

Neural networks

The best pattern recognition engines today

- ✓ Handles uncertainty

Symbolic AI

Historically well developed and has deep algorithmic understanding

- ✗ Does not handle uncertainty

Neural networks versus Symbolic AI

Neural networks

The best pattern recognition engines today

- ✓ Handles uncertainty
- ✓ Use raw data to improve behavior

Symbolic AI

Historically well developed and has deep algorithmic understanding

- ✗ Does not handle uncertainty
- ✗ Does not use raw data to improve behavior

Neural networks versus Symbolic AI

Neural networks

The best pattern recognition engines today

- ✓ Handles uncertainty
- ✓ Use raw data to improve behavior
- ✓ Easy to design

Symbolic AI

Historically well developed and has deep algorithmic understanding

- ✗ Does not handle uncertainty
- ✗ Does not use raw data to improve behavior
- ✗ Difficult to design and deploy

Neural networks versus Symbolic AI

Neural networks

The best pattern recognition engines today

- ✓ Handles uncertainty
- ✓ Use raw data to improve behavior
- ✓ Easy to design
- ✗ Do not easily work with easily stated rules about a problem

Symbolic AI

Historically well developed and has deep algorithmic understanding

- ✗ Does not handle uncertainty
- ✗ Does not use raw data to improve behavior
- ✗ Difficult to design and deploy
- ✓ Naturally defined to use declaratively stated rules

Neural networks versus Symbolic AI

Neural networks

The best pattern recognition engines today

- ✓ Handles uncertainty
- ✓ Use raw data to improve behavior
- ✓ Easy to design
- ✗ Do not easily work with easily stated rules about a problem
- ✗ Does not really perform reasoning (and planning, search, or any recursive algorithmic behavior)

Symbolic AI

Historically well developed and has deep algorithmic understanding

- ✗ Does not handle uncertainty
- ✗ Does not use raw data to improve behavior
- ✗ Difficult to design and deploy
- ✓ Naturally defined to use declaratively stated rules
- ✓ Well suited for reasoning (and planning, search, or any recursive algorithmic behavior)

Neural networks versus Symbolic AI

Neural networks

The best pattern recognition engines today

- ✓ Handles uncertainty
- ✓ Use raw data to improve behavior
- ✓ Easy to design
- ✗ Do not easily work with easily stated rules about a problem
- ✗ Does not really perform reasoning (and planning, search, or any recursive algorithmic behavior)
- ✗ Poor auditability of decision process

Symbolic AI

Historically well developed and has deep algorithmic understanding

- ✗ Does not handle uncertainty
- ✗ Does not use raw data to improve behavior
- ✗ Difficult to design and deploy
- ✓ Naturally defined to use declaratively stated rules
- ✓ Well suited for reasoning (and planning, search, or any recursive algorithmic behavior)
- ✓ Transparent decision process

Neural networks versus Symbolic AI

Neural networks

The best pattern recognition engines today

- ✓ Handles uncertainty
- ✓ Use raw data to improve behavior
- ✓ Easy to design
- ✗ Do not easily work with easily stated rules about a problem
- ✗ Does not really perform reasoning (and planning, search, or any recursive algorithmic behavior)
- ✗ Poor auditability of decision process

Symbolic AI

Historically well developed and has deep algorithmic understanding

- ✗ Does not handle uncertainty
- ✗ Does not use raw data to improve behavior
- ✗ Difficult to design and deploy
- ✓ Naturally defined to use declaratively stated rules
- ✓ Well suited for reasoning (and planning, search, or any recursive algorithmic behavior)
- ✓ Transparent decision process

What do we want: Best of both worlds

Neural networks versus Symbolic AI

Neural networks

The best pattern recognition engines today

- ✓ Handles uncertainty
- ✓ Use raw data to improve behavior
- ✓ Easy to design
- ✗ Do not easily work with easily stated rules about a problem
- ✗ Does not really perform reasoning (and planning, search, or any recursive algorithmic behavior)
- ✗ Poor auditability of decision process

Symbolic AI

Historically well developed and has deep algorithmic understanding

- ✗ Does not handle uncertainty
- ✗ Does not use raw data to improve behavior
- ✗ Difficult to design and deploy
- ✓ Naturally defined to use declaratively stated rules
- ✓ Well suited for reasoning (and planning, search, or any recursive algorithmic behavior)
- ✓ Transparent decision process

What do we want: Neuro-Symbolic Models

Neural network + symbolic programs

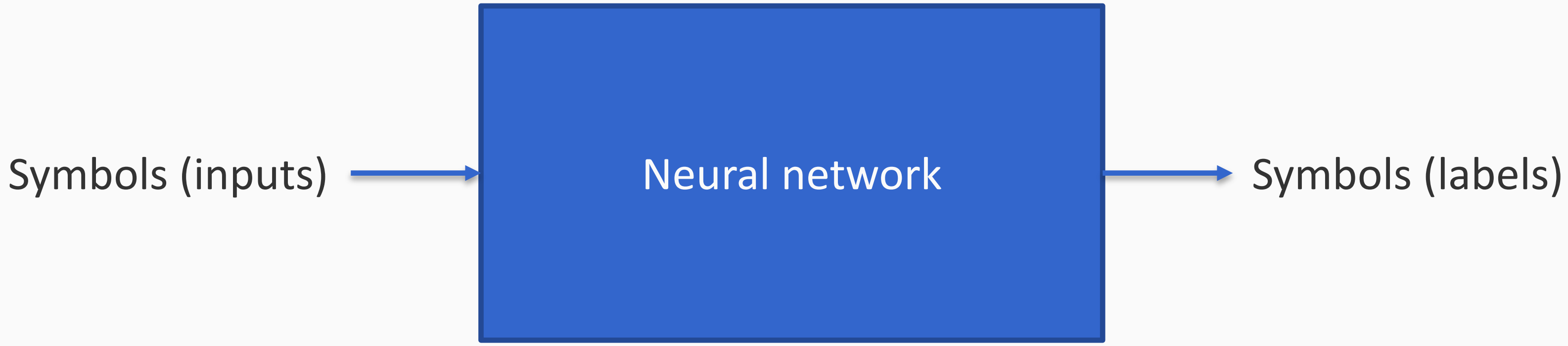
Neural networks involve vectors and differentiable functions

Rules are symbolic objects (i.e. consisting of discrete decisions)

How do they interface with each other?

We saw different kinds of neuro-symbolic interfaces

Approach 1: Just train the neural network



Convert symbolic inputs into vectors, operate with them and produce symbolic outputs

Standard neural networks do this

Approach 1: Just train the neural network

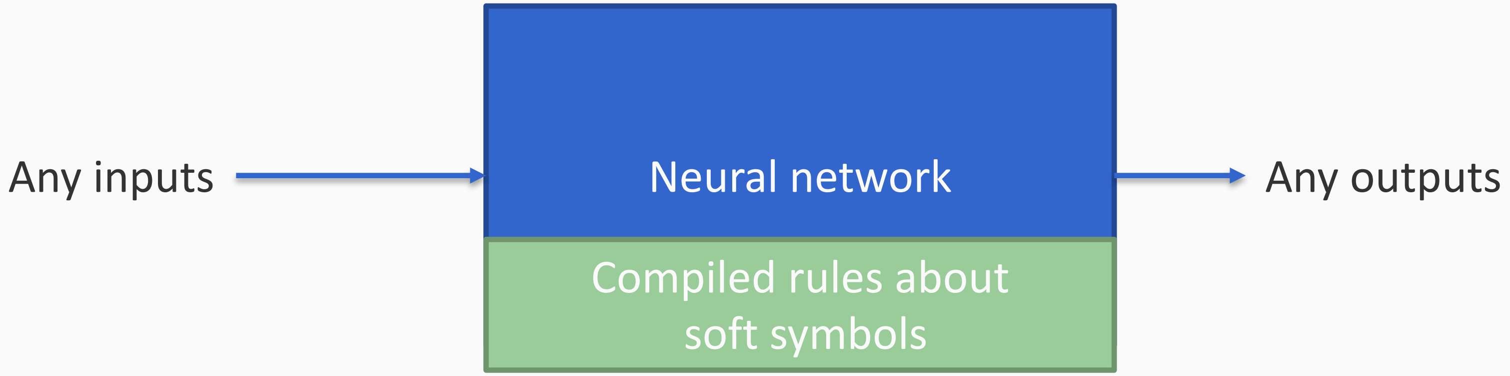
Suppose we have a task where we require that any model satisfies some properties P1, P2,…

Imagine that we have a *massive* amount of training data, covering a diverse set of domains for the task

Because the training data is correct, presumably the data satisfies the properties.

Can we just train a model to fit the data? In doing so, its predictions will satisfy the required properties because the data does

Approach 2: Logic to design networks

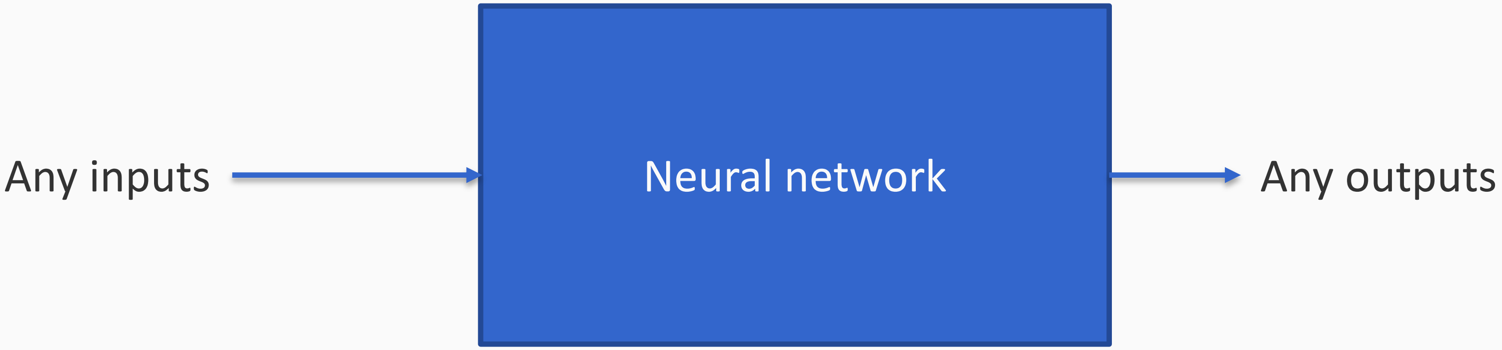


Treat some nodes within a neural network as soft symbols

Compile symbolic rules (i.e. constraints) into neural network submodules about these symbols to augment the network architecture

Key question: How can we re-architect the network so that the resulting architecture (by construction) satisfies the constraint? Or almost satisfies the constraint?

Approach 3: Logic to design loss functions



Training: data loss + compiled(rules)

Symbolic rules (about outputs or both inputs and outputs) compiled into a form that augments the training process

Implicit claim: Neural networks can internally represent the rules. *Is this valid?*

Approach 3: Logic to design loss functions

Most neural networks are opaque and the only interfaces we have are at the inputs and outputs

This means that most constraints will also about them

Can we write loss functions about the outputs that encourage the model to satisfy the constraints? Each constraint will be mapped to its own loss

We can then use any learning algorithm/optimizer

Approach 4a: Structured inference

Convert symbolic inputs into vectors, operate with distributions over labels

Apply probabilistic inference (with symbolic knowledge) over the top of network outputs to produce final outputs: *Structured prediction*

Approach 4a: Structured inference

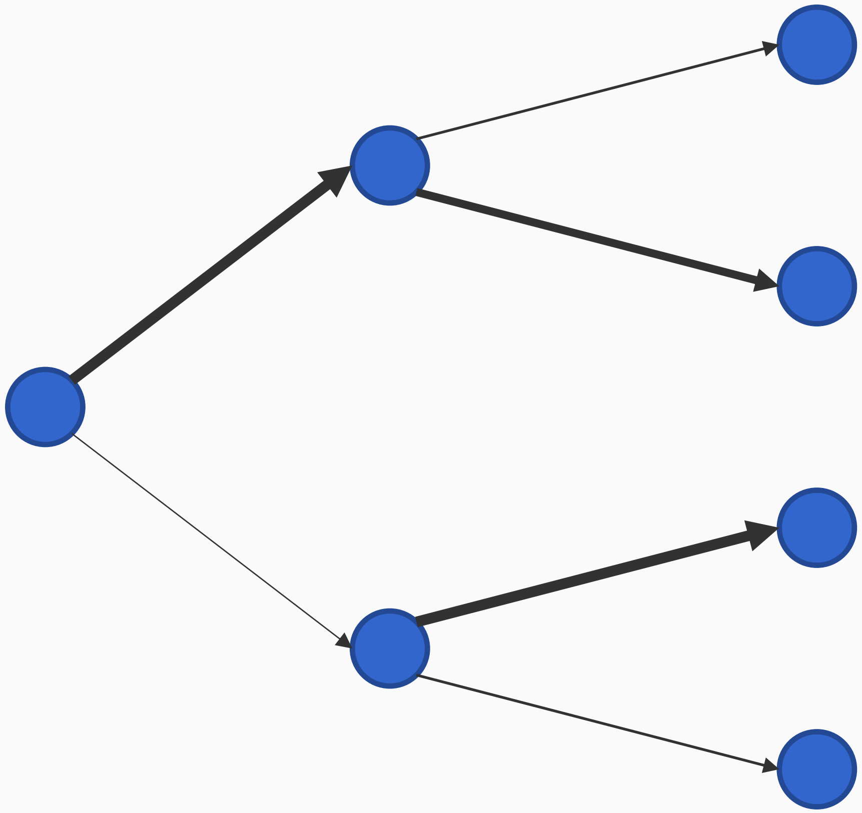
The neural network produces probabilities over its output space. Suppose our rules are constraints about the outputs

If our constraints are purely about outputs, then we could set up a constrained optimization problem whose solution is guaranteed to satisfy the constraints

The neural network creates the optimization problem for an input instance, the symbolic solver solves it

This could provide supervision at training time and/or can be used only at deployment

Approach 4b: Neural network guided symbolic problem solver



Use a neural network to guide a program that navigates a symbolic space by scoring paths (e.g. game playing, combinatorial inference, etc)

Approach 4b: Neural network guided symbolic problem solver

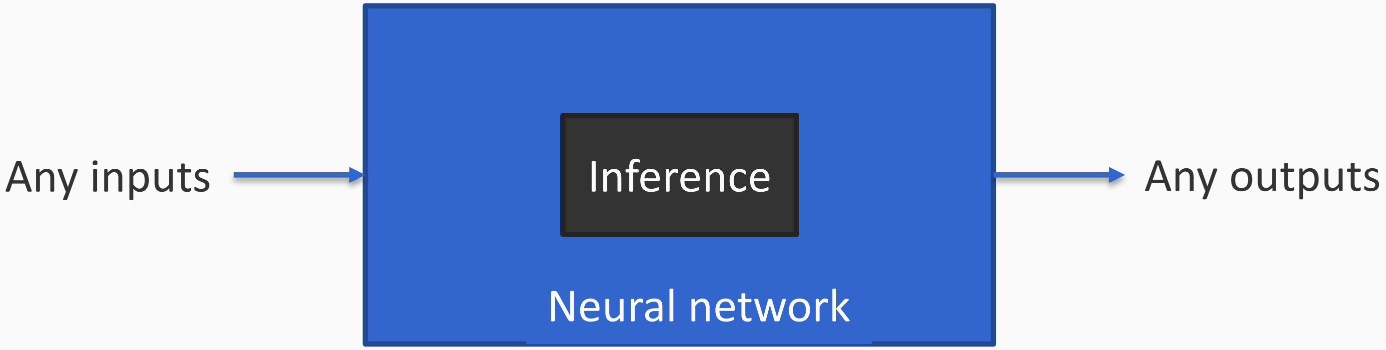
The neural network produces probabilities over its output space

The network probabilities could guide a search system that explores a richer (compositional) output space

The output space may involve symbolic constraints that the search system will incorporate

What are some difficulties with this strategy?

Approach 5: Symbolic programs within neural networks



Symbolic reasoning inside a neural network

Differentiability is an issue.

Approach 5: Symbolic programs within neural networks

Suppose we have a symbolic program that operates over the predictions of a model

Some Examples:

- A hard threshold node within a neural networks
- A sampling node within a network
- The program executes a python program created by a model
- The network calls a search engine to fetch documents on which the network continues working with
- The network informs a game engine to run a simulation, which in turn produces the output

Can the outputs of these programs provide supervision to train the network? Can the eventual deployment of the network involve such programs?

What we saw this semester

Semester overview

Part 1: The “logic-as-loss” approach

- The idea that logic can be compiled into loss functions
- Two high level strategies for “logic-as-loss”
  - Weighted model counting, semantic loss
  - Soft logic, t-norms
- Strengths and limitations of this strategy

Semester overview

Part 2: The “logic-as-network” approach

- The idea that logic can be compiled into neural networks
- Using soft logic for this
- Strengths and limitations of this strategy

Semester overview

Part 3: The “structured prediction” approach

- Structured prediction and training
- Network guided search (e.g. beam search)
- Integer programming based methods

Semester overview

Part 4: Symbols inside neural networks

– Straight through estimators

– Gumbel softmax

– The REINFORCE algorithm

There are more questions than answers in this field

Some open questions

- When will these ideas work? When will they fail?
- What kinds of assumptions are we making when we use these ideas?
- Can we build tools that implement these solution patterns we saw?
- Instead of logic, can we state preferences in language?
- How do we make the experimental parts more robust
- Can we apply these ideas for building explainable models?
- Can we apply these ideas for teaching models about fairness, transparency, accountability criteria?
- How can we apply these ideas to a new task X? Where does the knowledge come from?

There are more questions than answers in this field

Some open questions

- When will these ideas work? When will they fail?
- What kinds of assumptio making when we use the
- Can we build tools that in these solution patterns w
- Instead of logic, can we s preferences in language?

- How do we make the experimental parts more robust
- s for building
- s for teaching ransparency,
- How can we apply these ideas to a new task X? Where does the knowledge come from?

Neuro-Symbolic models: Challenges, but also opportunities

What can you do with them?