

Nicht in der Welt: Benötigen LLMs Erfahrungswissen?

Wolfgang Eppler, Reinhard Heil

- Ist Sprache unsere ganze Welt?
- Was bedeutet
 - „Die Grenzen meiner Sprache bedeuten die Grenzen meiner Welt.“ ?
- Ist die Unterscheidung “know how – know that“ sinnvoll?
- Bezieht sich Sprache auf etwas in der Welt?
- Auf was bezieht sich die Ausgabe eines LLMs?
- Wie machen LLMs Erfahrungen? Machen sie überhaupt welche?
- Wenn nein, was bedeutet das?

1 Grounding Problem of AI

Harnad 1990 “The Symbol Grounding Problem” of symbolic AI systems

Bender et al. 2021 “Stochastic Parrots”

Bender and Koller 2020 “... Meaning, Form, and Understanding in the Age of Data”

- Grounding problem persists on non-symbolic AI systems
 - Meaning is a relation between
 - form and communicative intent
 - form and something external to language
- LLMs have no meaning, only form
- LLMs miss reference on actual world

3 Philosophers on Empiricism

Quine 1951 “Two Dogmas of Empiricism”

- Analytic-synthetic distinction **is wrong**.
- Reductionism **is wrong**.

Davidson 1973 “On the Very Idea of a Conceptual Scheme” - third dogma of empiricism

- Dualism of scheme (form, language) and content (world, objects) **is wrong**.

→ Empiricism is false.

Reference and truth are problematic.

- No reference from linguistic terms to immediate experience
- No relation to something external to language

→ No grounding of language?

4 Grounding by Practices

Grounding

- not by reference or relation

Grounding

- by handling objects
- by experiential learning during practical action

Brandom, Habermas as two representatives of pragmatism



Generated by DALL-E

5 Brandom: Parrot and Observer

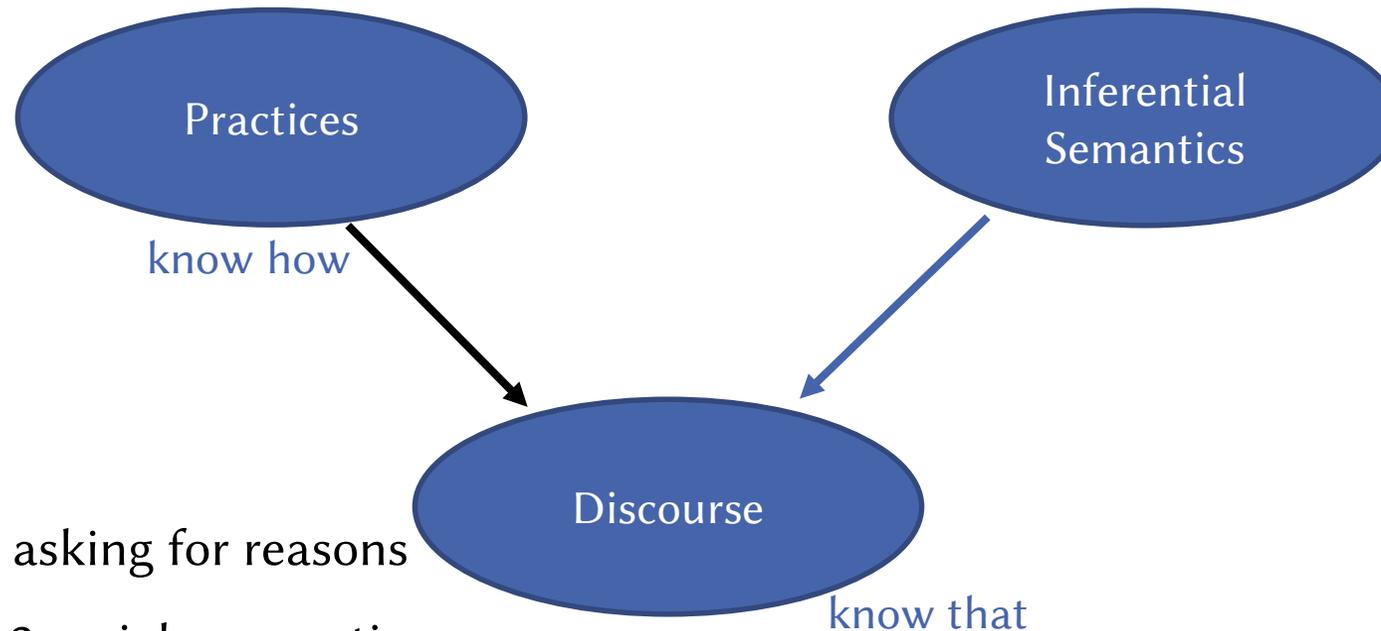
- Brandom: Difference between parrot and observer
 - Parrot responds with „That's red“ seeing a red thing.
 - Observer understands the concept of „red“
 - Understand a concept:
 - Know material inferences of a concept
 - i.e. know premises and conclusions of concept
 - „If I get a red-wrapped present, it must be my birthday.“
- Inferential Semantics (Sellars 1953 „Inference and Meaning“)
- Do LLMs have inferential semantics?
 - Do LLMs have concepts?

→ Yes!



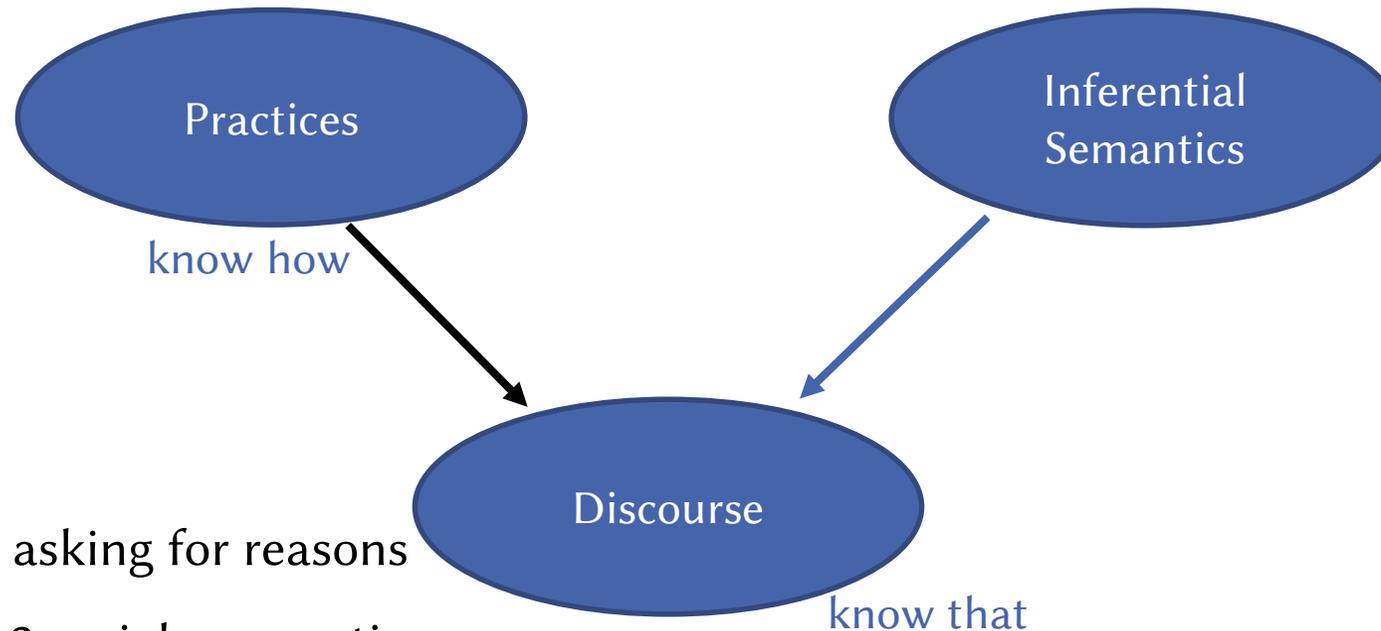
Generated by DALL-E

5 Brandom in „Making it Explicit“ 1996



- Game of giving and asking for reasons
- Commitments with 2 social perspectives
 - undertaken by Speaker
 - attributed by Hearer to Speaker
- Keep score of commitments → memorize and update them
David Lewis 1978 “Scorekeeping in a Language Game“

5 Brandom in „Making it Explicit“ 1996



- Game of giving and asking for reasons
- Commitments with 2 social perspectives
 - undertaken by Speaker
 - attributed by Hearer to Speaker
- Keep score of commitments → memorize and update them
David Lewis 1978 “Scorekeeping in a Language Game“

→ LLM

- no scorekeeping
- no memorization of commitments

6 Habermas (Theorie des kommunikativen Handelns 1981)

Criticizes functionalist reason

- distinction of life-world and system

Life-world

- individuals interact and communicate
- engage with an objective world
- share understandings, cultural traditions, and social norms

System

- refers to the structured, organized aspects of society
- operates through formal mechanisms and instrumental rationality
- communication and social integration of life-world is disrupted

6 Habermas (Wahrheit und Rechtfertigung 1999)

Life-world

- Transition from action to discourse
 - **If** “practices damage and contradictions emerge in life-world”
then subjective certainties **evolve to** “claimed truths”
- Discourse: justification by exchange of arguments
 - Participants adopt a reflexive attitude with two different perspectives
 - Participants argue about the truth of statements

Learning processes

- with actions in life-world
- with discourse in life-world

→ LLM

- no discourse on true arguments,
- no learning process in the life-world

7 What about Large Language Models (LLM)?

Hypothesis: „LLMs are not grounded in our actual world.“ ✓

- No incremental, no continual learning
- No learning by individual experience
- No parameter update during dialog
- No practice in life-world

due to batch processing of pretraining and finetuning

8 Folgerungen für unsere Lebens- und Arbeitswelt

Fehlende Verankerung der LLMs in der Welt

- Unzuverlässigkeit, Unwahrheit, Unmöglichkeit, mit LLMs ernsthaft zu diskutieren und sie dadurch zu verbessern
- Schwurbeleien
- amüsant beim Spielen und Ausprobieren
- lästig und schädlich bei der Arbeit – aber nicht immer!

Folgen abhängig von Anwendungsgebiet

- Leistungsbewertung oder Rekrutierung neuer Mitarbeitenden
- Zusammenfassende Berichte für Entscheidungsträger, z.B. in der Technikfolgenabschätzung
- Dienstleistungssoftware für Mitarbeitende

8 Folgerungen für unsere Lebens- und Arbeitswelt

Folgen bei Leistungsbewertung oder Rekrutierung neuer Mitarbeitenden

- Algorithmische Auswertung von zig-tausend Parametern pro Bewerbungsgespräch
- unwahrscheinlich, dass Fehler der KI bemerkt werden (wenn Ergebnis nicht eklatant falsch)

→ Keine Transparenz bei LLMs, keine Nachvollziehbarkeit

→ Ungerecht behandelte BewerberInnen sind ohnmächtig gegenüber sie betreffende Entscheidungen.

High risk nach EU AI Act

- Einstellung und Auswahl von Arbeitnehmern
- Leistung und Verhalten von Arbeitnehmern
- KI-Ergebnisse sollen nachvollziehbar sein

→ Nur unter gewissen Umständen erlaubt, aber nicht verboten

8 Folgerungen für unsere Lebens- und Arbeitswelt

Folgen bei Zusammenfassung von Berichten für Entscheidungsträger, z.B. in der Technikfolgenabschätzung

- Von LLMs produzierte Berichte klingen plausibel
- Berichte sind in kürzester Zeit fertig
- Berichte geben nur sprachlich fixiertes Wissen wieder
- nicht kodifizierte Erfahrung spielt keine Rolle

→ aber: Erfahrungswissen ist für Bewertung und Prognosen unersetzlich

- LLMs unzuverlässiger je aktueller eine Debatte
 - Siehe Einschätzung von Covid während der Covid-19-Phase
 - Siehe Fragen nach „gestohlenen“ Wählerstimmen nach der US-Präsidentschaftswahl 2020

→ Als Hilfsmittel nützlich

→ Aber große Vorsicht!

8 Folgerungen für unsere Lebens- und Arbeitswelt

Folgen bei Dienstleistungssoftware für Mitarbeitende

- immer wieder Punkte, an denen es nicht weitergeht (weil Position nicht vorschriftsmäßig ausgefüllt)
- Frustration, Hilflosigkeit, Inaktivität wegen Eigenheiten und (nichtmenschlichen) Fehlern
 - aufgrund maschinell ausgestellter Schreiben (von Behörden oder von Inkassounternehmen) oder
 - der Blockade elektronischer Formulare
- wird durch KI fortgesetzt, vermutlich mit Durchdringung unseres gesamten Alltags/ Arbeitstags
- Schnelligkeit und Bequemlichkeit contra Zuverlässigkeit

- Bei Dysfunktion kein Mensch mehr in der Kette
- Pfadabhängigkeit!
- Kein Zurück mehr!

8 Folgerungen für unsere Lebens- und Arbeitswelt

Trotz Annäherung der KI an Lebenswelt durch Fine-Tuning:

- KI fehlt Umgang mit der alltäglichen Lebenswelt (raum-zeitliche und sprachliche Welt mit Erfahrungen).
- KI versagt bei unseren selbstverständlichsten Handlungen.

→ Die menschliche Arbeits- und Lebensweise kommt in diesen Momenten ins Stocken.

- Der Mensch kennt das von kleineren und größeren Widersprüchen, die im Umgang mit der natürlichen und sozialen Welt auftreten.
- Der Mensch lernt aus diesen Erfahrungen.
- In Zukunft wird er aus den Widersprüchen mit der KI seine Erfahrungen machen.

→ Das ist nicht das, was die Menschheit und die Welt am dringendsten braucht.

- Die Anpassung unserer menschlichen Lebens- und Arbeitsweise an die Natur und die soziale internationale Gemeinschaft erfordert viel Gespür und Einfühlungsvermögen.
- Gerade da kann eine KI am wenigsten helfen.

Danke!

eppler@kit.edu

reinhard.heil@kit.edu