Impossibility of Unambiguous Communication as a Source of Failure in AI Systems

William J. Howe
Dr. Roman V. Yampolskiy
NLP Based AI Systems
Humans have intuitive understanding of each level of linguistic analysis, but current NLP systems do not. As a result, the many ambiguities in language can produce failure modes for AI systems at each of these levels.

---

"...an end to end approach will solve self-driving cars" ²

"...an autoregressive language model with 175 billion parameters, 10x more than any previous non-sparse language model"³

²https://twitter.com/comma_ai/status/1333902872421908480
³Brown, Tom B., et al. "Language models are few-shot learners."
End-to-End Approach

1. Benchmark performance improves as we move from hand-crafted features to end-to-end systems, sacrificing some explainability.

2. But our AI systems also should have human like understanding at each of the linguistic levels of analysis.

3. Current end-to-end approaches don’t achieve this multi-level understanding and explainability.
The Problem

1. Because current AI systems don’t have NLU (natural language understanding), there are inherent risks in controlling them with natural language.

2. At the same time, an AI system with NLU can be used to manipulate and influence using natural language ambiguity.
(1) Eliminate the target with a bomb.
Syntactic Ambiguity
Precisely the type of symbolic reasoning these NLP systems aren’t capable of.
Coding Schemes

One fix for syntactic ambiguity might be to use specially designed coding schemes like *prefix codes*.

<table>
<thead>
<tr>
<th>C(a)</th>
<th>C(b)</th>
<th>C(c)</th>
<th>abc to</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>01</td>
<td>00</td>
<td>10100</td>
</tr>
</tbody>
</table>

But this only addresses **one type of ambiguity**.

---

Phonological Ambiguity

There already exist real world speech recognition assistants and so there is the problem of *phonological ambiguity* too:

(2)  a. the tail of the dog...
    b. the tale of the dog
AI Exploiting Ambiguity

Conversely, an AI with natural language understanding can effectively be used to:

1. Generate ambiguous news headlines
2. Generate legal documents, software requirements, and contracts with ambiguity.
3. Deployed as a malicious web agent
Safety Concerns

1. In the near term, increased explainability and interpretability will increase confidence in AI systems.

2. Neuro-symbolic AI approaches will help bridge the gap towards NLU.

3. Increased regulation to help curb bad actors and rampant web agents or bots with NLU capabilities.