Human-level play in the game of *Diplomacy* by combining language models with strategic reasoning

Meta Fundamental AI Research Diplomacy Team (FAIR)

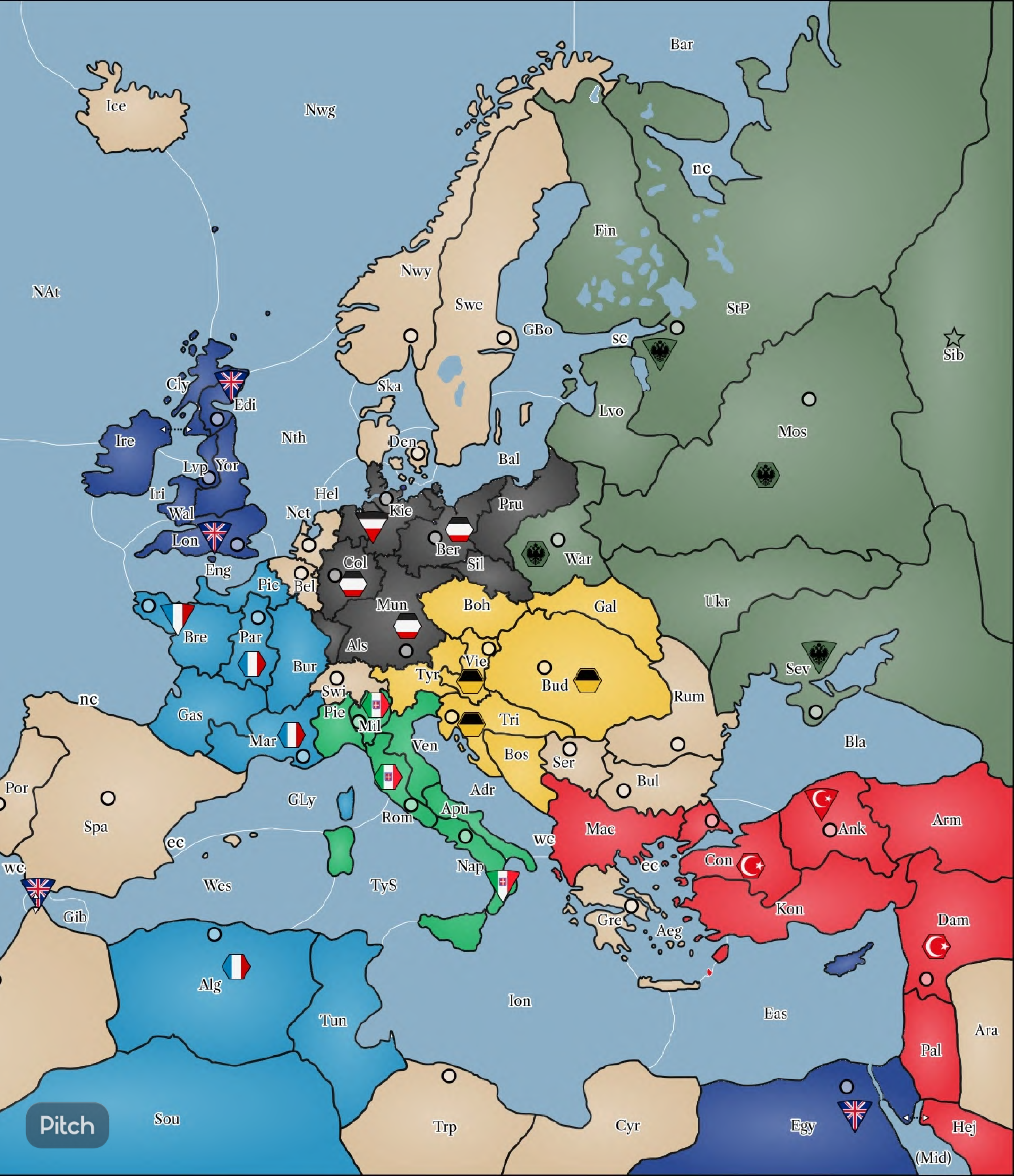
Ward Pennink

2023-02-17

Agenda

- **Diplomacy**
  - Game
  - Challenges
- **CICERO**
  - Dialogue module
  - Strategic Reasoning module
- Performance
- Final remarks

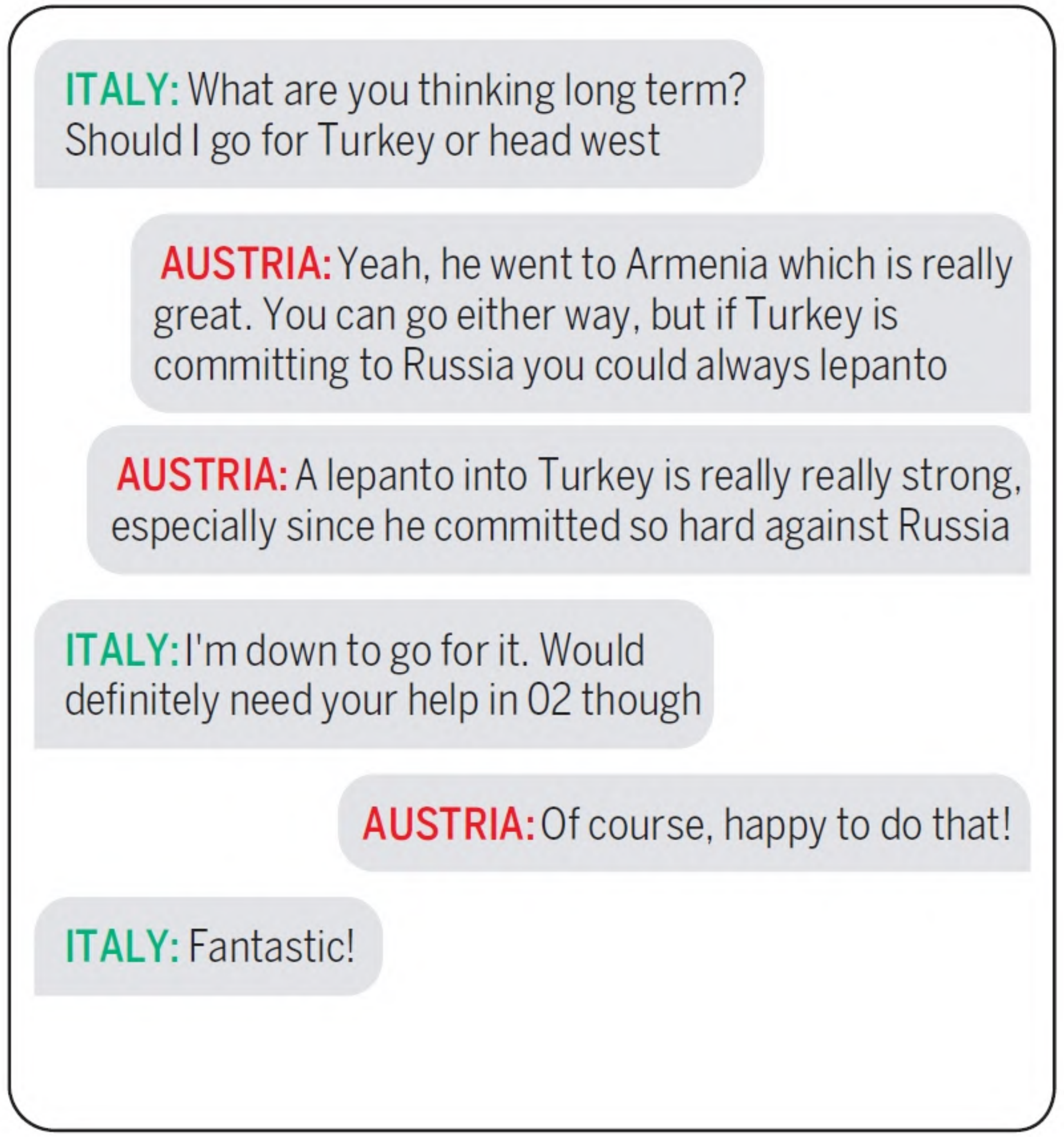
DIPLOMACY



Diplomacy

- 7 players
- **Goal** - gain control of the map
- Simultaneous move game
- Basic conquer mechanics increase importance of **coordination**
  - Private dialogue between 2 players

Dialogue examples in Diplomacy



ITALY: What are you thinking long term? Should I go for Turkey or head west

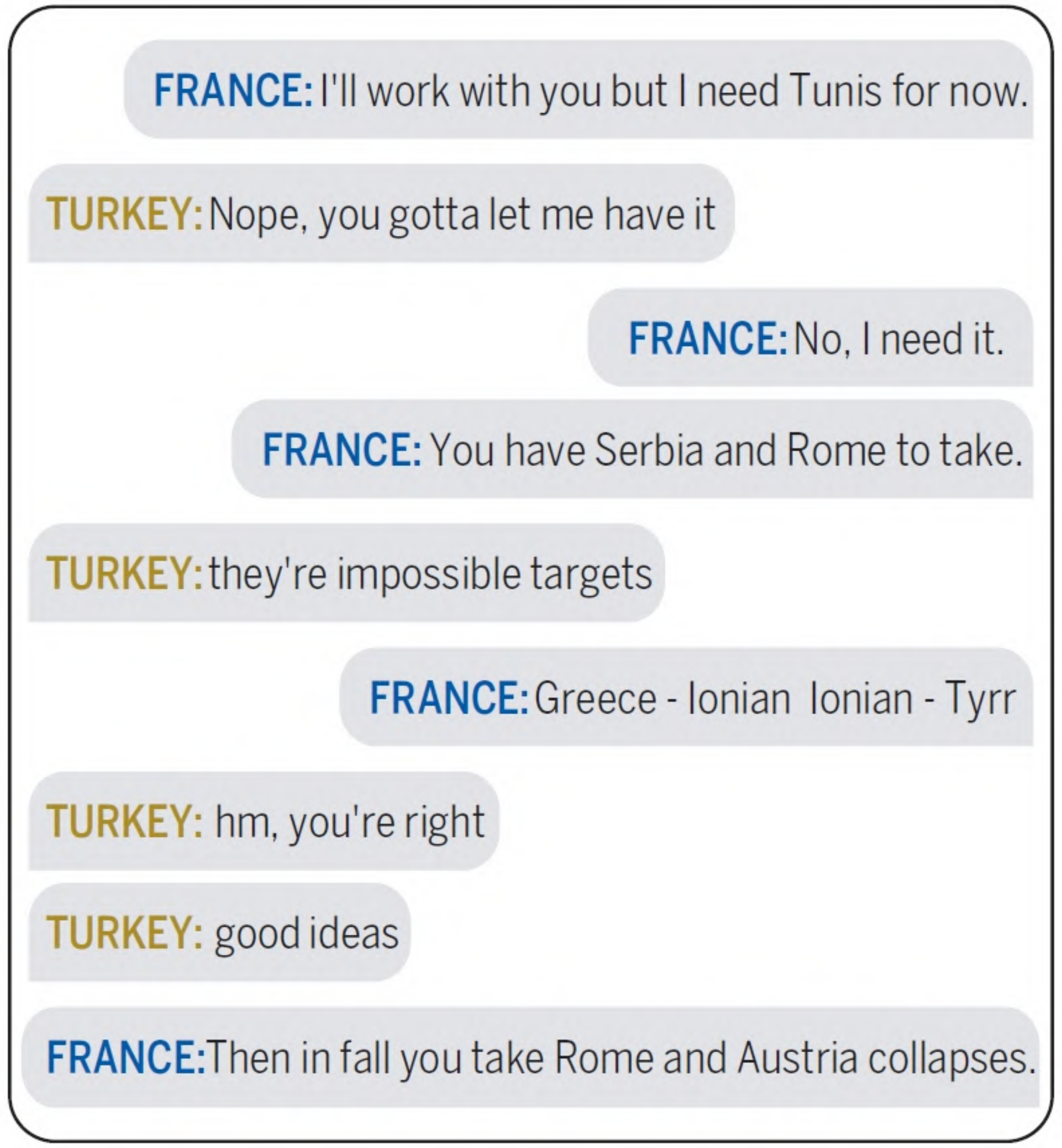
AUSTRIA: Yeah, he went to Armenia which is really great. You can go either way, but if Turkey is committing to Russia you could always lepanto

AUSTRIA: A lepanto into Turkey is really really strong, especially since he committed so hard against Russia

ITALY: I'm down to go for it. Would definitely need your help in 02 though

AUSTRIA: Of course, happy to do that!

ITALY: Fantastic!



FRANCE: I'll work with you but I need Tunis for now.

TURKEY: Nope, you gotta let me have it

FRANCE: No, I need it.

FRANCE: You have Serbia and Rome to take.

TURKEY: they're impossible targets

FRANCE: Greece - Ionian Ionian - Tyrr

TURKEY: hm, you're right

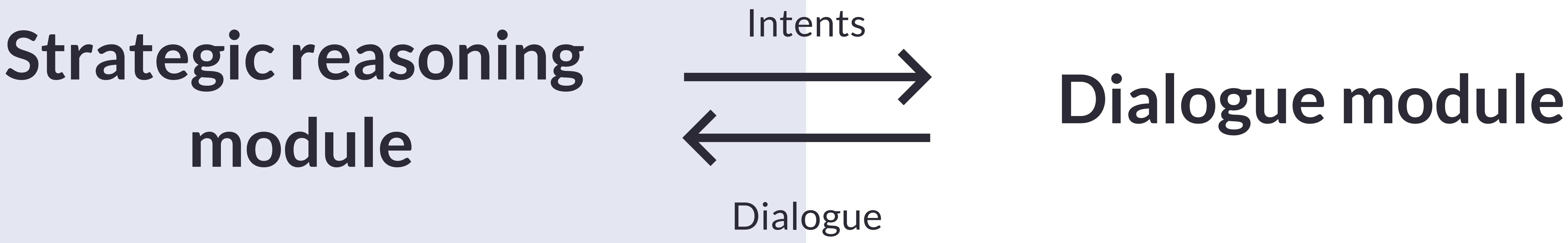
TURKEY: good ideas

FRANCE:Then in fall you take Rome and Austria collapses.

Challenges in Diplomacy

- **Self-play** Reinforcement Learning incompatible with human-play
- **Multi-agent** setting
- Messages must be **grounded** in dialogue history, game state, and goals.
- Success requires building **trust** with other players

CICERO



Strategic reasoning module

Intents

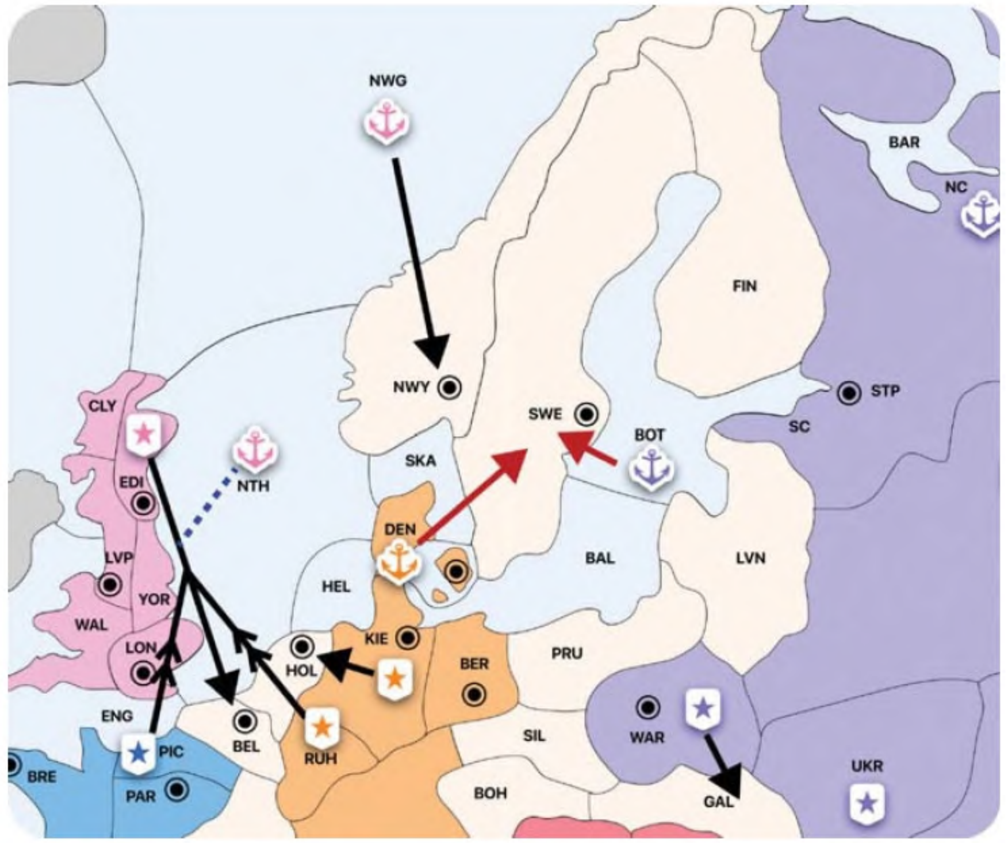
Dialogue

Dialogue module

Dialogue module

- Supervised training of Transformer model on Diplomacy messages
- Language model conditioned on **intents**:
  - *A message has intent **z** if **z** is the most likely set of actions that the sender and recipient will take.*
- Advantages of conditioning on intents:
  - Automatically captures legal and strategic moves
  - Provides an interface between the Strategic module and the Dialogue module

England convoys an army to Belgium with the support of France while taking Norway in a manner friendly to Russia

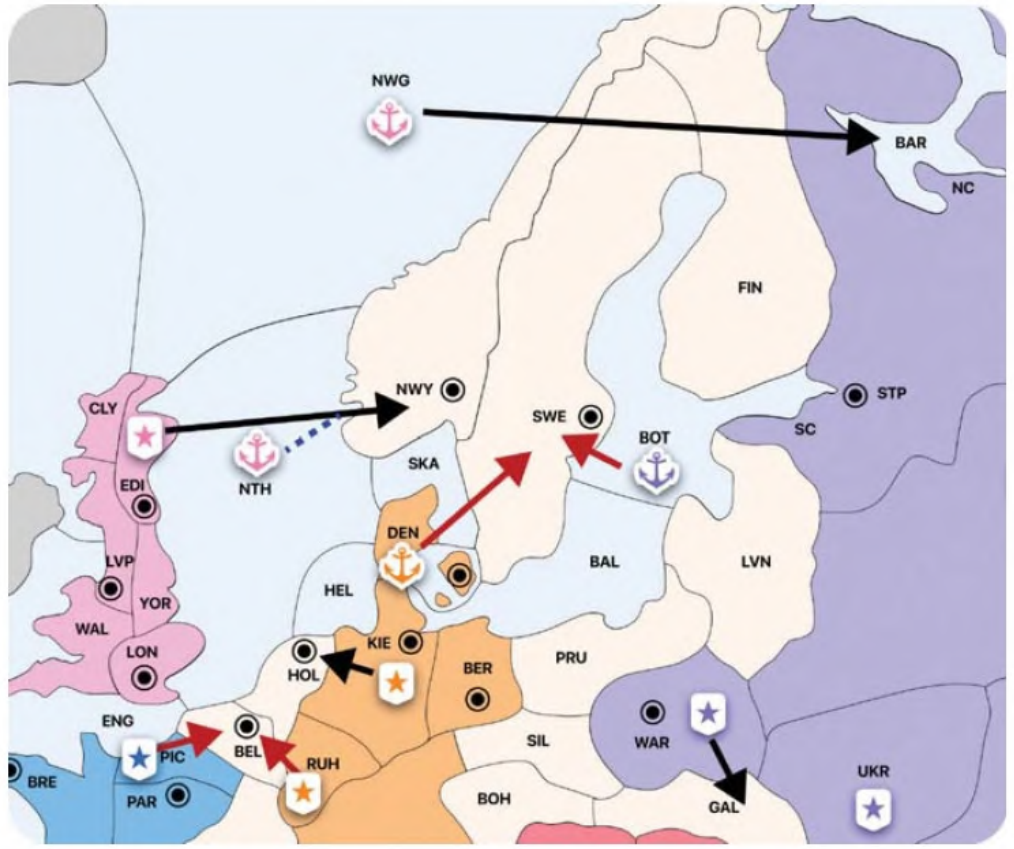


ENG → FRA Mind supporting Edi - Bel?

ENG → GER Do you wanna support my convoy to Bel? With Italy going aggressive France will fall quickly and we can make gains off of both Russia and France.

ENG → RUS How are you thinking Germany is gonna open? I may have a shot at Belgium, but I'd need your help into Den next year.

England convoys an army to Norway and also aggressively moves to the Barent Sea against Russia

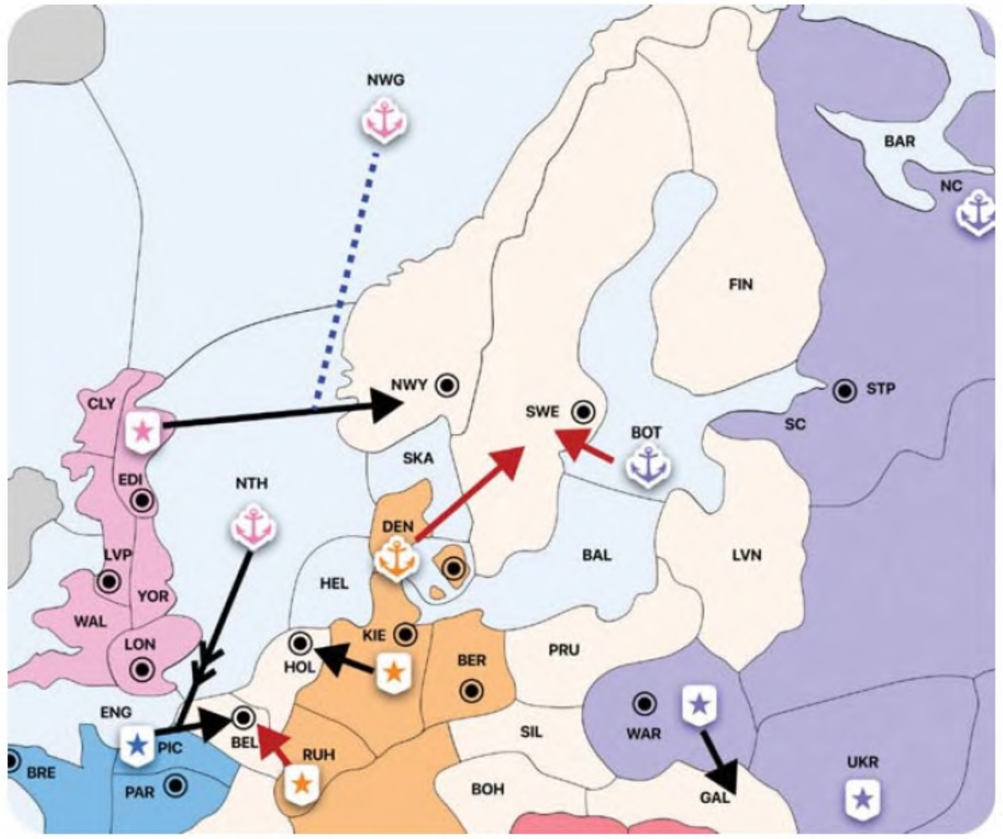


ENG → FRA What's the plan for Bel?

ENG → GER I will be convoying my army to Norway and moving to Barrents Sea.

ENG → RUS Hey, I moved to Edi to keep my options open. Do you have a problem with me putting an army in Nwy?

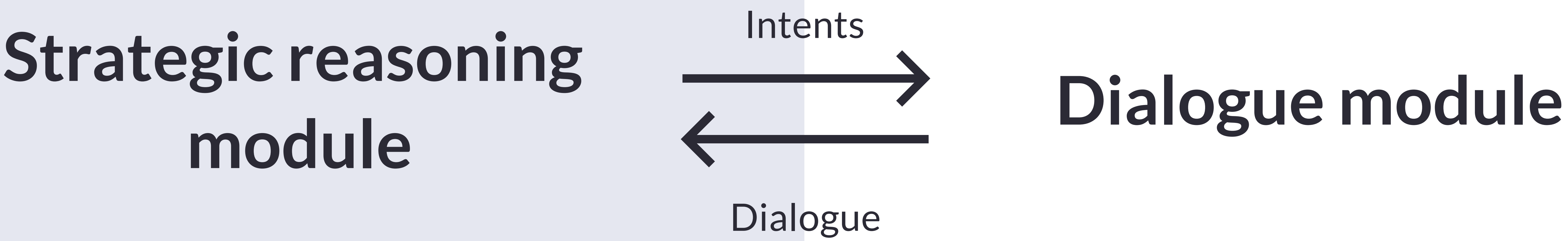
England convoys an army to Norway and supports France to gain Belgium, preveting Germany from taking Belgium



ENG → FRA Let me know if you want me to support you to Bel, otherwise I'll probably poke Hol.

ENG → GER Looks like you'll get three builds unless France bounces you! Are you gonna bounce Russia or not?

ENG → RUS Hey, I moved to Edi to keep my options open. Do you have a problem with me putting an army in Nwy?



Strategic reasoning module

Intents

Dialogue

Dialogue module

Strategic Reasoning Module (1/2)

Goal: predict other players' strategies for the current turn according to the state of the board and the shared dialogue → choose own strategy

Modelling other players' strategies:

- Anchor policy τ_i based on supervised learning from human data
- **piKL**
  - Each turn treated as its own subgame with simultaneous moves
  - Assumes each player seeks a strategy π_i that maximizes their own utility function,
  - While minimizing KL divergence between π_i and τ_i

$$U_i(\pi_i, \pi_{-i}) = u_i(\pi_i, \pi_{-i}) - \lambda D_{KL}(\pi_i || \tau_i)$$

- $u_i(\pi_i, \pi_{-i})$ calculated based on RL learning with self-play

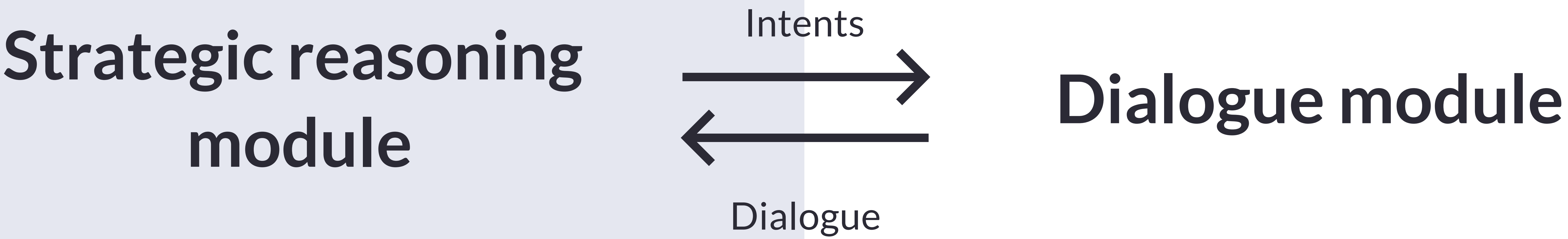
Strategic Reasoning Module (2/2)

Goal: predict other players' strategies for the current turn according to the state of the board and the shared dialogue → choose own strategy

Modelling own strategy:

- CICERO chooses action a_i that best responds to other players' joint strategy:

$$\arg\max_{a_i} u_i(a_i, \pi_{-i}) + \lambda \log \tau_i(a_i)$$

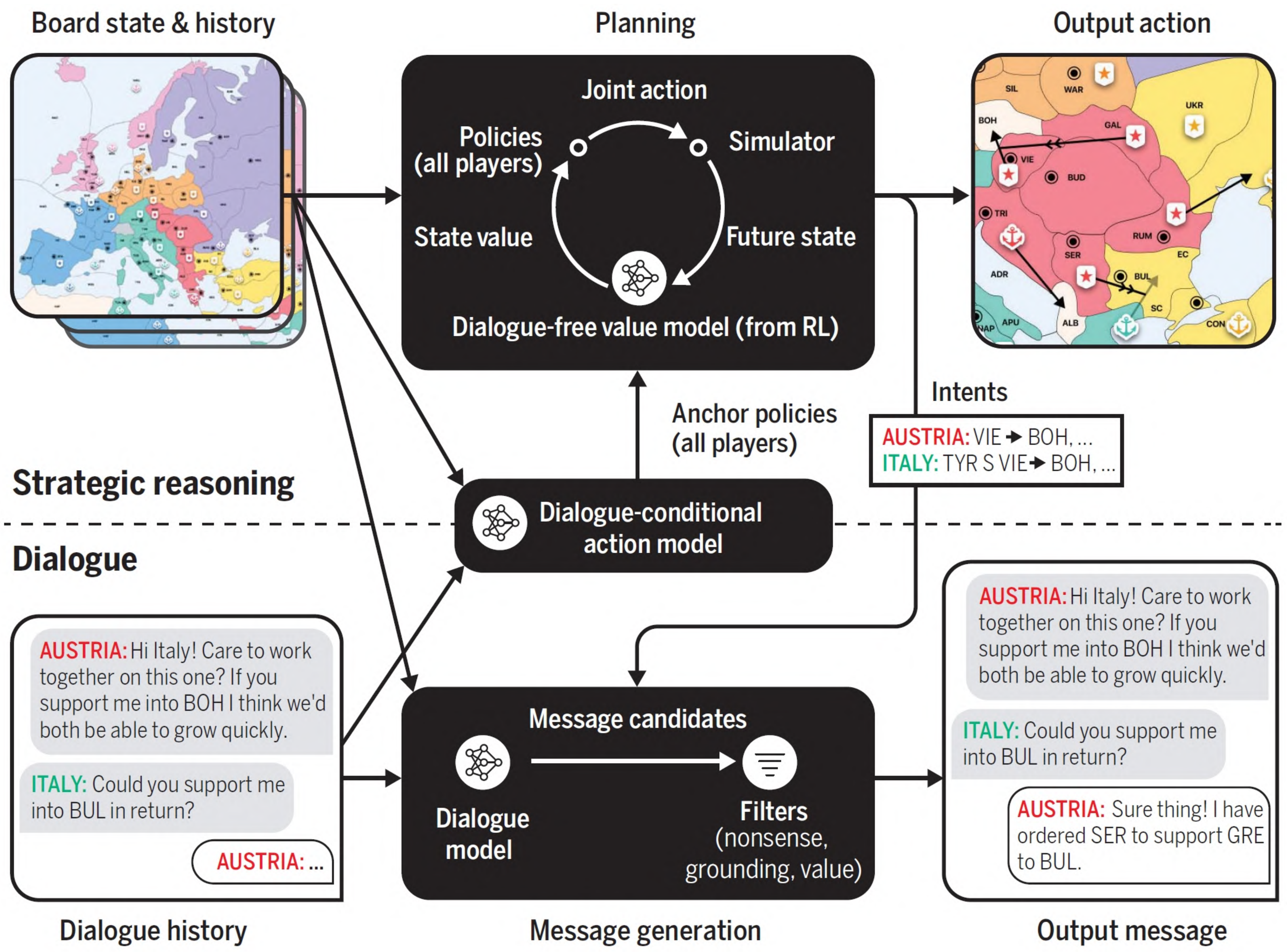


Strategic reasoning module

Intents

Dialogue

Dialogue module



Board state & history

Planning

Output action

Joint action

Policies (all players)

Simulator

State value

Future state

Dialogue-free value model (from RL)

Intents

AUSTRIA: VIE → BOH, ...
ITALY: TYR S VIE → BOH, ...

Anchor policies (all players)

Strategic reasoning

Dialogue-conditional action model

Dialogue

AUSTRIA: Hi Italy! Care to work together on this one? If you support me into BOH I think we'd both be able to grow quickly.

ITALY: Could you support me into BUL in return?

AUSTRIA: ...

Message candidates

Dialogue model

Filters (nonsense, grounding, value)

AUSTRIA: Hi Italy! Care to work together on this one? If you support me into BOH I think we'd both be able to grow quickly.

ITALY: Could you support me into BUL in return?

AUSTRIA: Sure thing! I have ordered SER to support GRE to BUL.

Dialogue history

Message generation

Output message

CICERO architecture

Performance

- 40 games of 5-minute negotiation time per turn
  - **Top 10%** of participants that played >1 game
  - **2nd** out of 19 participants that played >5 games
  - Mean score of **25.8%** compared to **12.4%** average score

Final remarks

Limitations

- 5-min negotiation limit makes for less complex dialogue
- CICERO is honest in its messages → problem for repeated play
- Occasional errors in messages

Conclusion

- CICERO

References

Meta Fundamental AI Research Diplomacy Team (FAIR)†, Bakhtin, A., Brown, N., Dinan, E., Farina, G., Flaherty, C., ... & Zijlstra, M. (2022). Human-level play in the game of Diplomacy by combining language models with strategic reasoning. *Science*, *378*(6624), 1067-1074.

Bakhtin, A., Wu, D. J., Lerer, A., Gray, J., Jacob, A. P., Farina, G., ... & Brown, N. (2022). Mastering the Game of No-Press Diplomacy via Human-Regularized Reinforcement Learning and Planning. arXiv preprint arXiv:2210.05492.

De Jonge, D., Baarslag, T., Aydoğan, R., Jonker, C., Fujita, K., & Ito, T. (2019). The challenge of negotiation in the game of diplomacy. In Agreement Technologies: 6th International Conference, AT 2018, Bergen, Norway, December 6-7, 2018, Revised Selected Papers 6 (pp. 100-114). Springer International Publishing.