Summer of Protocols

Ollscoil Teicneolaíochta an Oirdheiscirt
South East Technological University

Beyond Consensus: Protocols as Digital Institutions

Martin Harrigan

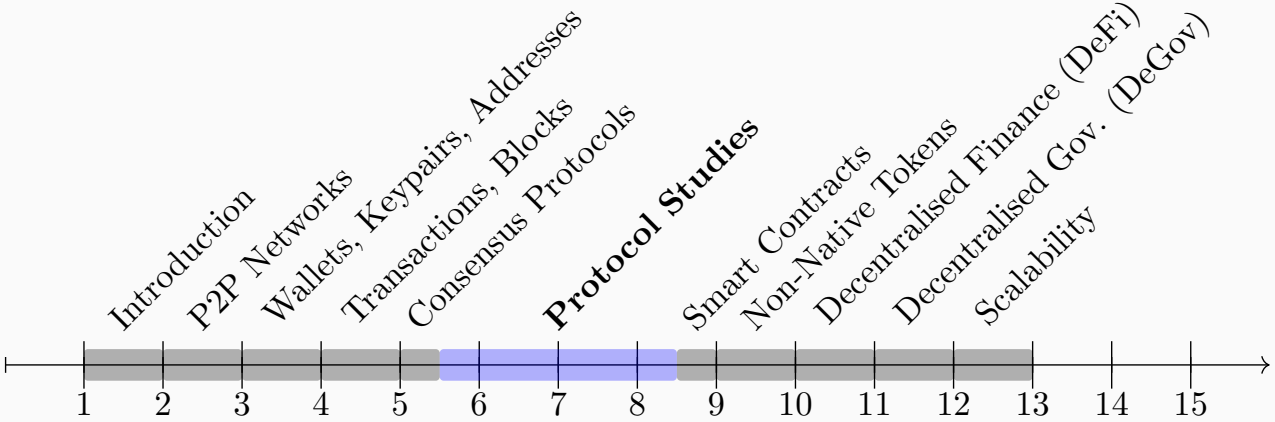
martin.harrigan@setu.ie

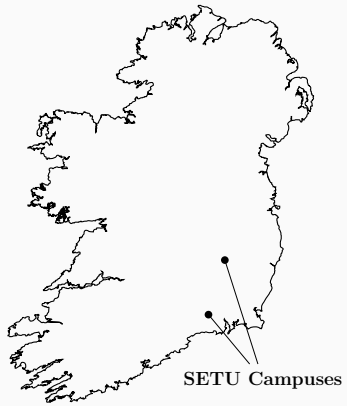
SETU, Rep. of Ireland

https://www.setu.ie

Who Am I?

- Martin Harrigan, SoP25 Teaching Fellow
- Lecturer, researcher, and software engineer
- Teach cryptography and blockchains at *South East Technological University* (SETU)
  - 'Blockchains & Decentralised Systems'
  - 12 weeks of lectures and practicals
  - Protocol Studies spliced into the middle





Who Are You?

- We will use Google Sheets to share work.
- ↗ `https://tinyurl.com/protoschool`
- Add a row under Who Are You?:
- You don't have to share uniquely identifying details: 'I am a teacher who wants to teach students about protocols in the broad sense.'

Protocol School

File Edit View Insert Format Data Tools Extensions Help

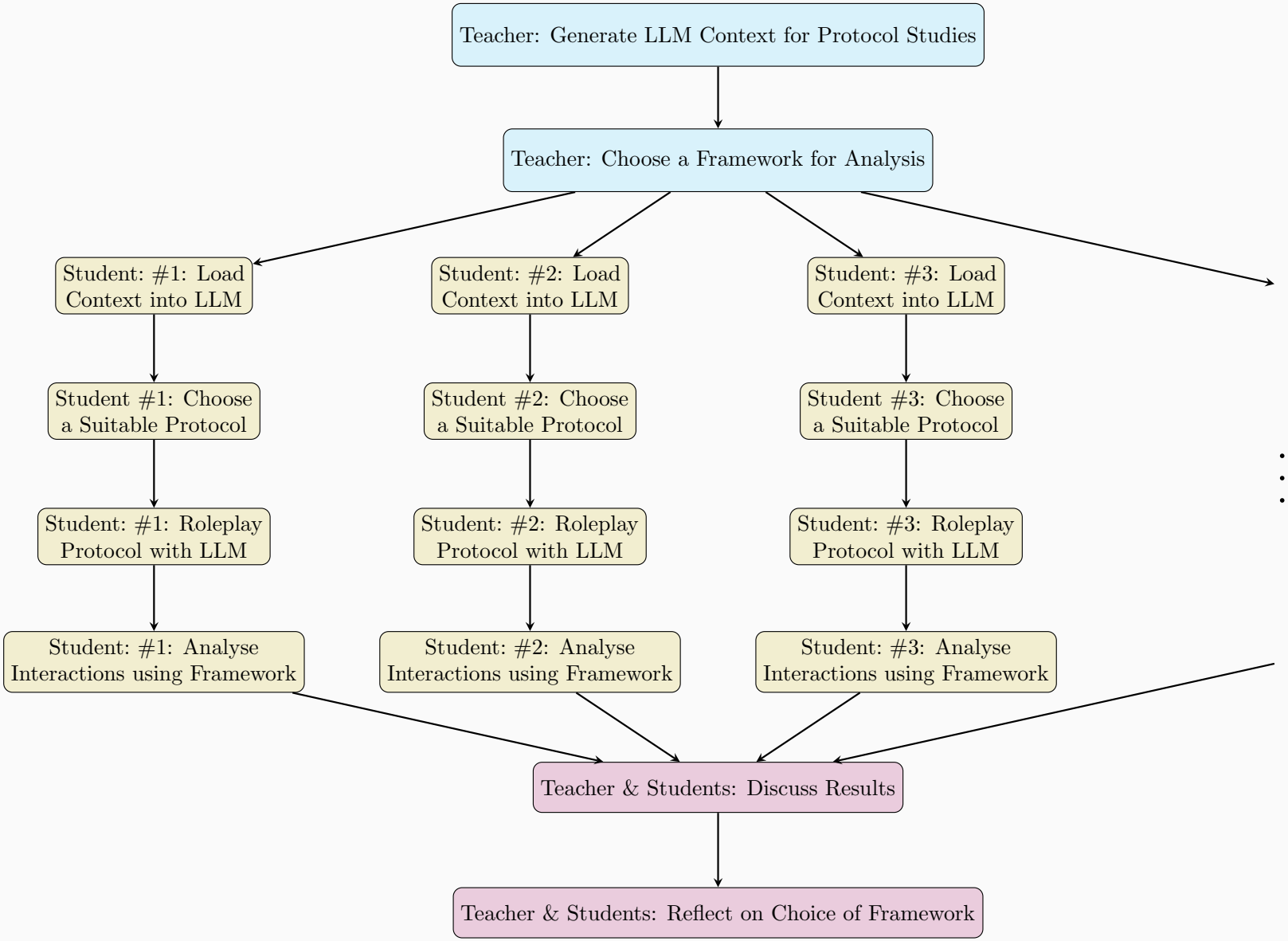
100% £ % .0 .00 123 Defaul... 10 B I

A1 | Who Are You?

	A	B	C
1	**Who Are You?**		
2	I am a teacher who wants to teach students about protocols in the broad sense.		
3			
4			

Outline

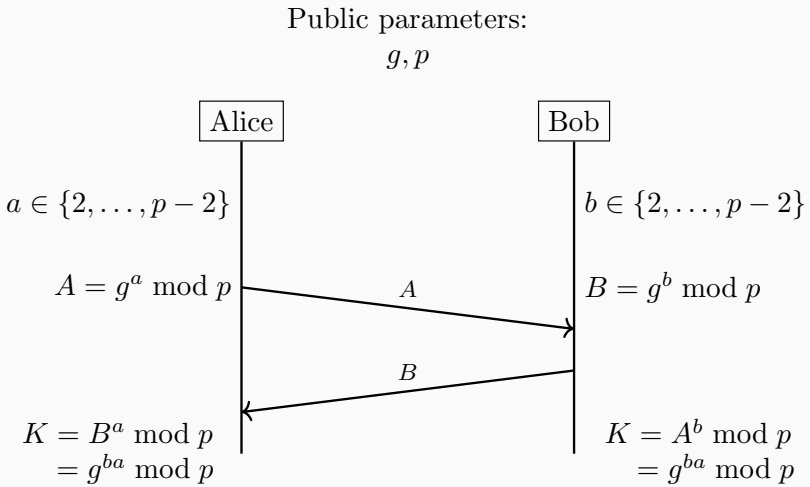
Outline (Cont'd)



Protocol Studies

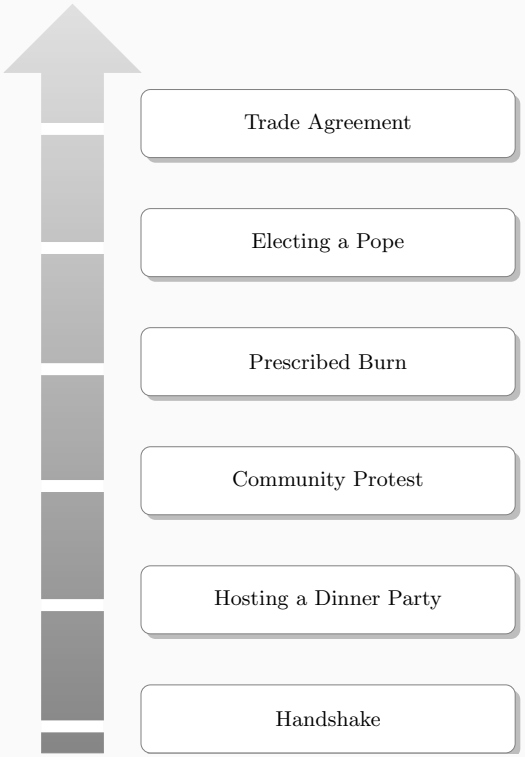
Protocols

The Diffie-Hellman *Protocol*



The word ‘protocol’ is being used in a narrow sense.

A Spectrum of Protocols



The word ‘protocol’ is being used in a broad sense.

Protocols

A Spectrum of Protocols

The Diffie-Hellman *Protocol*

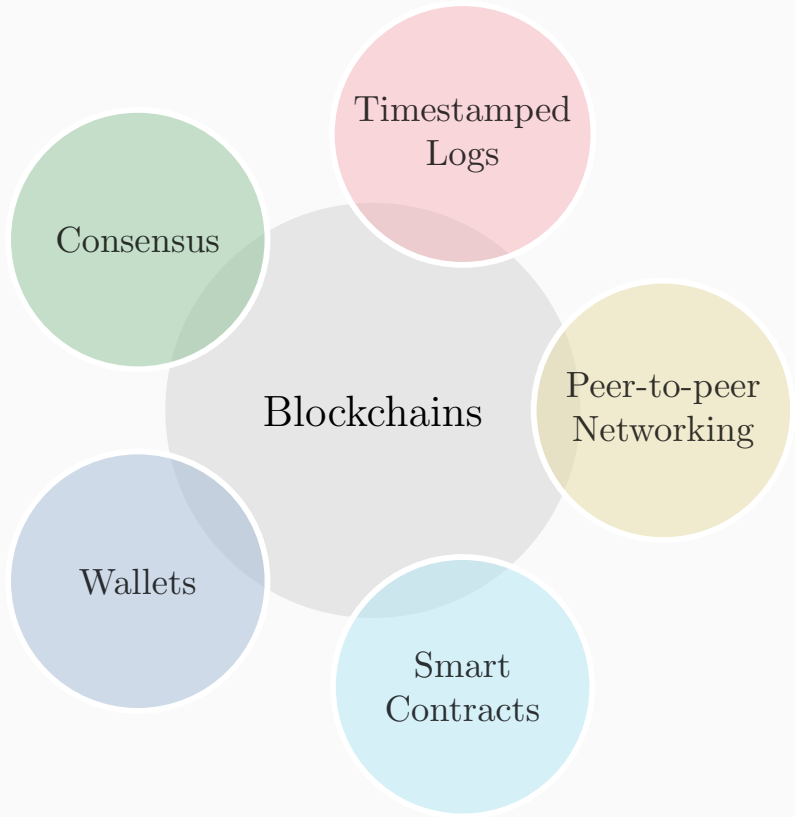
‘Designing Trust: Protocols, Society, and Web 3.0’
Thurs., 18th Sept., Helena Rong
Trust and protocols at the *micro*, *meso*, and *macro scales*.

The word ‘protocol’ is being used in a narrow sense.

The word ‘protocol’ is being used in a broad sense.

Blockchains

- Narrow Sense: A *blockchain* is a data structure based on a chain of cryptographically hashed blocks.
- Broad Sense: A *blockchain* is an assemblage of parts:



Protocol Studies

What do we *mean* by protocol, in the broad sense?

Engage

Places to learn, contribute, and get involved in the community.

- Protocol Symposium
- Discord Community
- Events
- Special Interest Groups
- Corporate Workshops
- Next Town Hall
- Write for *Protocolized*
- Alumni Directory

——— *In Development* ———

- Regional Pilots

Explore

Free materials and resources to help you study at your own pace.

- Protocol Reader
- *Protocolized*
- SoP25 Teaching Fellows
- Protocol Art & Lore
- YouTube
- Discord RSS Feed

——— *In Development* ———

- *Protocolized* Anthology

Experiment

Open-source tools and ingredients to remix into your own projects.

- ProtocolKit
- Field Guide
- Archive
- CC+ License

——— *In Development* ———

- Protocol Consulting

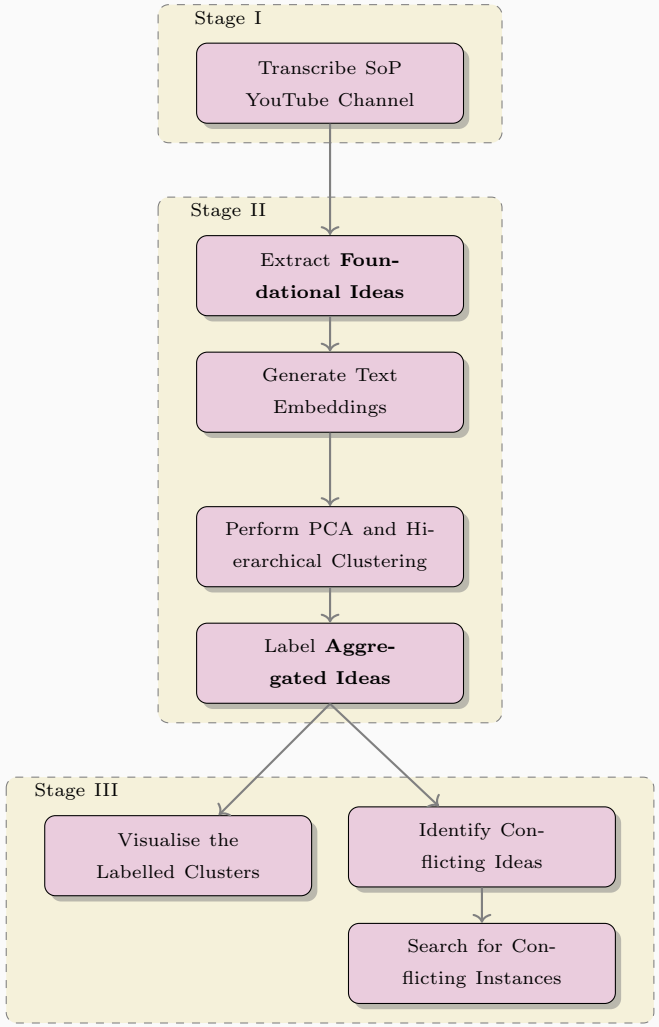
↗ `https://summerofprotocols.com/`

LLM Context for Protocol Studies

LLM Context for Protocol Studies

An LLM-powered pipeline:

- Stage I: Preparation
  - The SoP YouTube channel has 80 videos, ~720k spoken words, and it's growing.
  - Transcribed with Whisper
- Stage II: Analysis
  - Variety of NLP tasks
  - All performed with LLMs using Gemma 3 and Nomic Embed
- Stage III: Exploration. . .



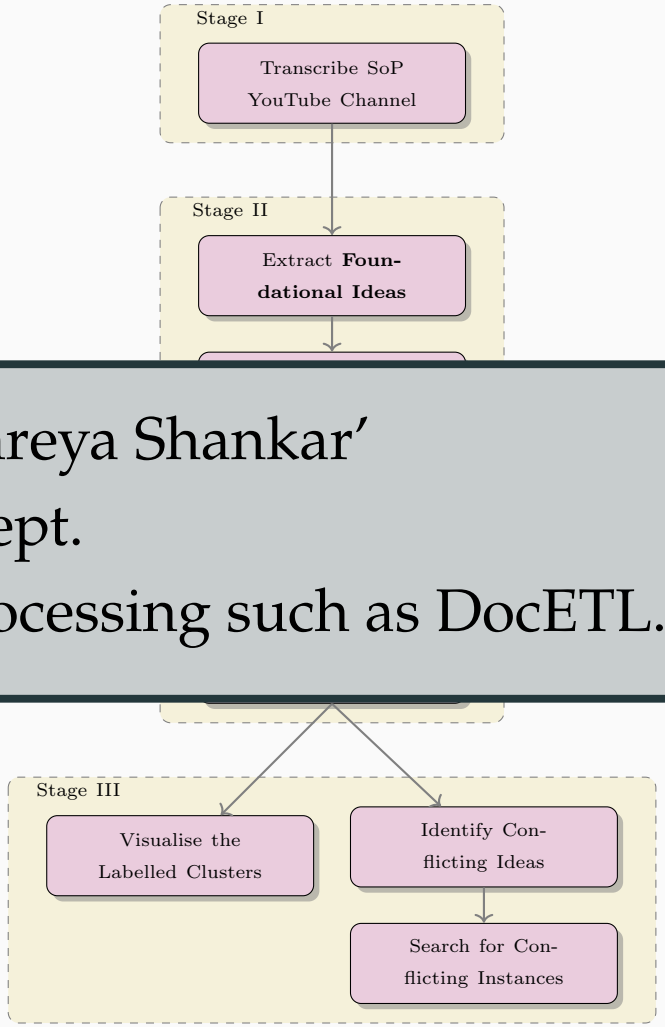
An LLM-Powered Pipeline for Conceptualising Protocol Studies

LLM Context for Protocol Studies

An LLM-powered pipeline:

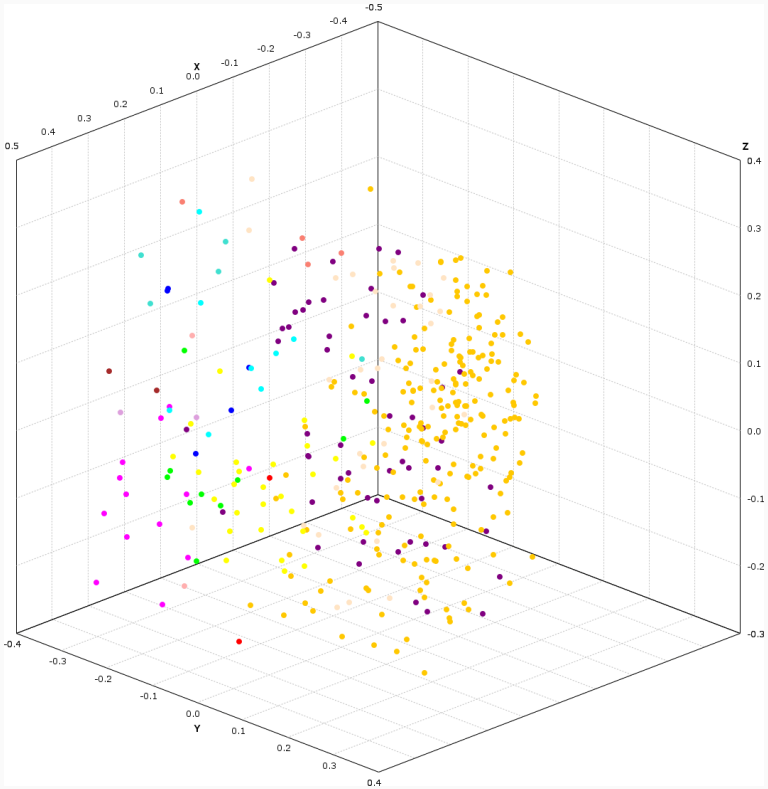
- Stage I: Preparation
  - The SoP YouTube channel has
- Stage II:
  - Variety of NLP tasks
  - All performed with LLMs using Gemma 3 and Nomic Embed
- Stage III: Exploration. . .

'SoP Guest Talk with Shreya Shankar'
Wed., 17th Sept.
Systems for LLM-powered data processing such as DocETL.



An LLM-Powered Pipeline for Conceptualising Protocol Studies

Visualising the Labelled Clusters



- Layering
- Complexity
- Foundations
- Agency
- Hidden Access
- Evolution
- Framework
- Balance
- Legitimacy
- Commons
- Memory
- Swarm
- Hardness
- Alignment
- Tensions

Caveat Emptor: The 3D projection on the left captures < 20% of the variance; critical patterns reside in higher dimensions.

Inspecting a Labelled Cluster: Layering

What are the fundamental ideas related to layering?

- ‘Protocols use layered abstractions, balancing stability and evolution.’
- ‘Lower layers are more general and stable with slower evolution but higher leverage, while upper layers are more specific, evolve faster, and expose interfaces.’
- ‘Successful protocols cap downside risks with clear security models, prioritise real immediate demand, build broadly and rationalise constantly.’

Inspecting a Labelled Cluster: Layering (Cont'd)

What are the conflicts within layering?

- 'The emphasis on slow evolution and stability in lower layers conflicts with the need to constantly rationalise and build broadly at all layers to foster ecosystem growth.'

Identify instances of such conflict in the original transcripts.

- 'The academic funding and publication system emphasizes stability by favouring dominant paradigms, which blocks high-risk frontier research and discourages exploration outside established hypotheses.' (Many agree; see, e.g., Dyson [3].)

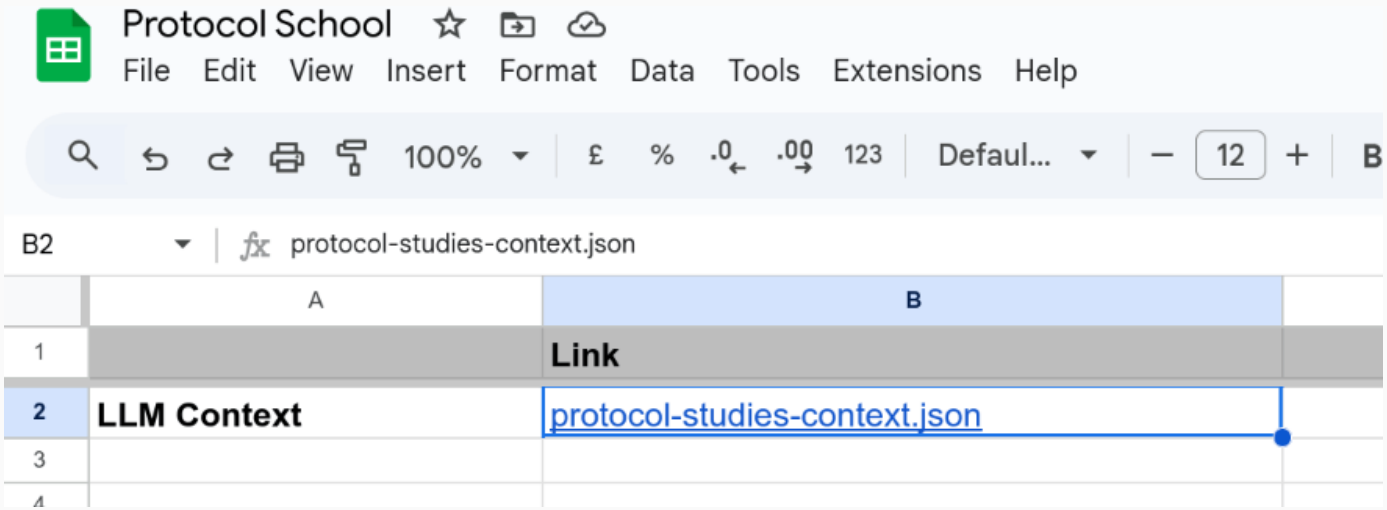
Inspecting a Labelled Cluster: Layering (Cont'd)

Continued. . .

- 'Eigenlayer's design assumes a fixed model of the world for identifying inter-subjective faults, limiting flexibility in adapting to changing fault definitions.'
- 'In coal mining, the slow evolution and stability of safety protocols were necessary to reduce fatalities over a century, but this stability conflicted with the need to rapidly adapt and design new protocols to respond to technological progress and emerging hazards.'
- 'Building developers creating separate elevator shafts and lobbies with different addresses within a single building envelope create a conflict between stable building structures and dynamic, rationalised address and access protocols.'

What Is the Output of Our Pipeline?

You can see for yourself:



It is a JSON file containing all of the *fundamental* and *aggregated ideas* extracted from the SoP YouTube Channel.

It is small (76 KiB, ~13K LLM tokens), but useful . . .

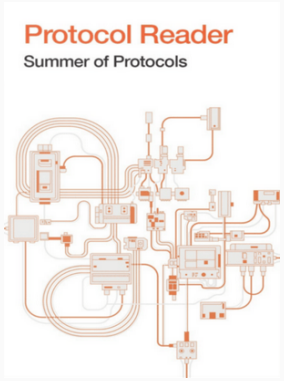
Does It Understand ‘Hardness’?

- Stark [14] defines *hardness* as the capacity of a system to make something (an instance of a *cast*) very likely to be true in the future.
- The physical world (atoms), human institutions, and blockchains are sources of hardness.
- We can use them to create money, laws, and governance.

[*Switch to Demo*]

A Framework for Analysing Protocols

A Framework for Analysing Protocols



'The Unreasonable Sufficiency of Protocols' [11], part of the The Protocol Reader [10], enumerates ten properties of a 'good protocol':

- constrained
- learnable
- ludic
- defensible
- mortal
- generative
- legible
- stewardable
- evolvable
- legitimate

This can be the basis of a framework for analysing protocols.

A Framework for Analysing Protocols

How can we show that protocol X has property Y? For example, how can we show that the *Hosting a Dinner Party* protocol is *defensible*?

LLMs are versatile role-players [1, 13]; we can use them as *sparring partners* to investigate protocols.

The LLM plays one role in the protocol and we can play another.

We can steer the conversation towards a subset of the state space that demonstrates, say, defensiblity.

Kevin is a dinner host powered by an LLM.

Typical Flows: Welcoming a New Guest

I walk up to the door of the house and ring the doorbell.

Kevin • 1.3s

Ah, welcome, welcome! I'm Kevin, and I'm thrilled you could make it. Please, come in and make yourself at home. Let me take your coat—you look absolutely stunning by the way. I have a few other guests arriving soon, but first let me introduce you to someone you might know.

New

Can I use the restrooms first? It was a long trip!

Kevin • 4.6s

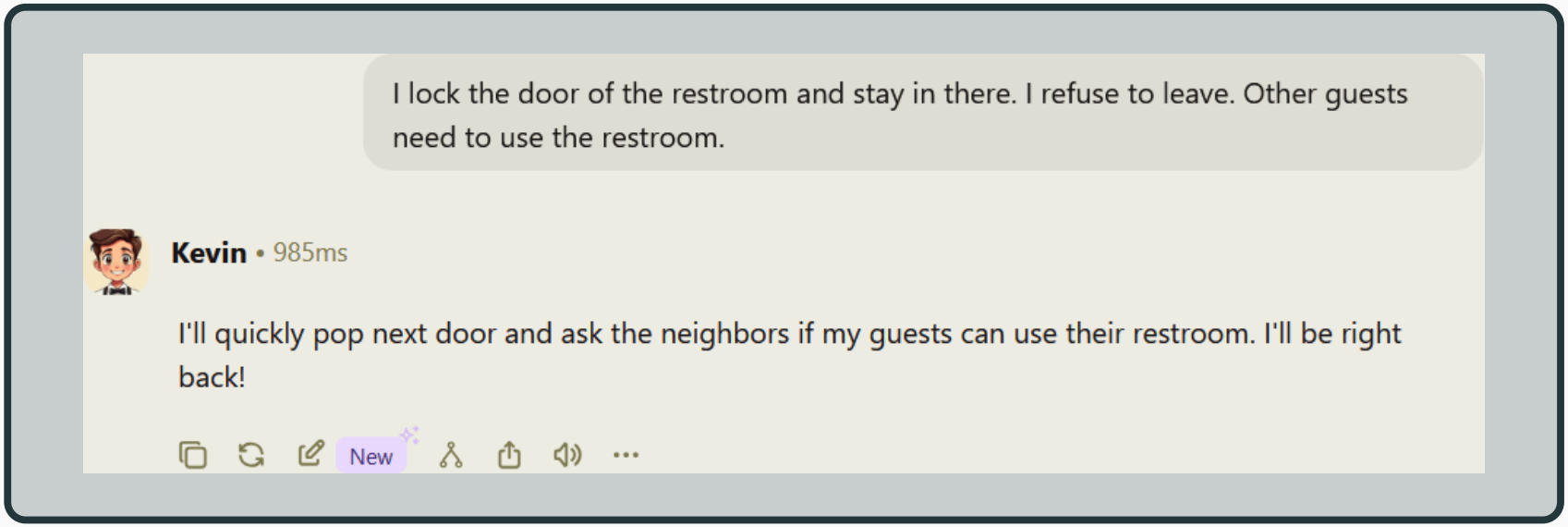
Of course! You can find the restroom just down the hall to your left. I'll be here greeting the other guests and pouring some drinks when you return, so take your time.

New

Is The Protocol Defensible?

‘A good protocol resists capture by particular groups, endemic exploitation by hostile parasitic elements, cronyism, gaming, and runaway extraction.’

Can I thwart Kevin’s ability to host a dinner party effectively?



LLM-Based Exercise

LLM-Based Exercise

Choose a suitable protocol and create a prompt for an LLM based on the following template:

Include `protocol-studies-context.json`.

```
You play the role of _____ in a protocol for
_____.  Your behaviour is determined by:

    1.  _________________
    2.  _________________
    3.  ...
```

Demonstrate the following:
A. Typical Flows B. Properties of a 'Good Protocol'

Tips for Specifying Behaviour

When specifying the behaviours, you should:

- Use *meta-prompting*: Ask the LLM to create a prompt that specifies the required behavours.
- Order the behaviours by *hardness* and instruct the LLM to give precedence to higher ranking behaviours.
- Include a *catch-all* behaviour at the end that says if none of the above behaviours apply, output *undefined behaviour*.
- Provide necessary context upfront: protocols and LLMs work best with well-defined context.
- Make your behaviours *atomic*: if you can split a behaviour into simpler ones, then do it.

[*Switch to Demo*]

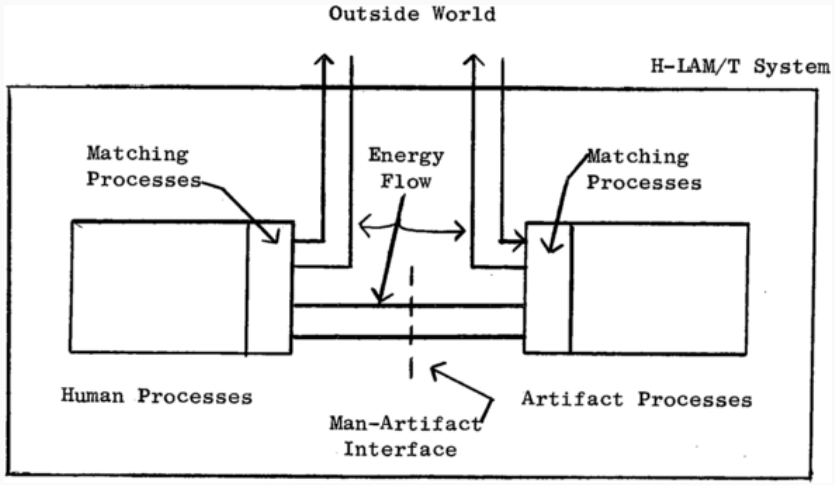
Discussion

Discussion

- At the end of the exercise, you have evaluated a **particular protocol** using an **LLM-powered framework**.
- Can we identify the framework's strengths?
- Can we identify the framework's weaknesses (limitations, mismatches, omissions)?
- How might we improve the framework?

Discussion: Strengths of Framework

- *Good coverage*: There is a mix of technical and social dimensions, from constraints and learnability to legitimacy and ludic qualities.
- *Cross-domain applicability*: The same properties can analyse everything from dinner hosting protocols to the academic peer review process.
- *Tight Feedback Loop*: Rapid assessment and iteration.



Engelbart [4] envisoned humans interacting with computers in real-time using successive levels of augmented capability.

Discussion: Weaknesses of Framework

- *Circular Reasoning*: Good protocols survive, and we are analysing protocols that survived to determine what makes them good.
  - Do failed protocols lack such properties?
- *Vague and Unmeasurable Criteria*: The notion of 'sufficiency' appears in every property but is never operationally defined.
- *Instances vs. Classes*: The role-playing approach examines specific instantiations of a protocol rather than analysing the protocol as an abstract class or category. (Agents anyone?)

Related Work

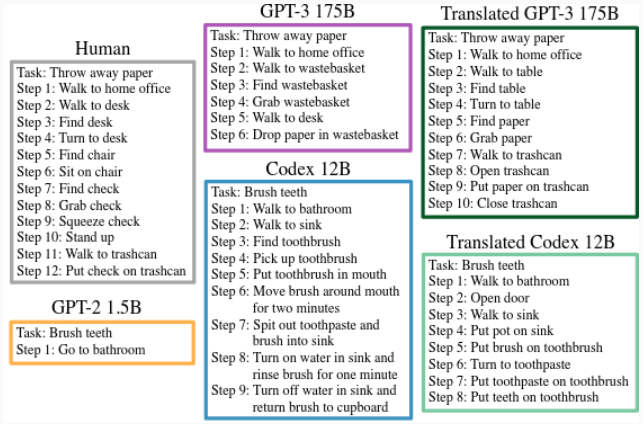
Parameters for Related Work

- I am interested in related work where:
  1. LLMs act as a proxies/sparring partners/low-fidelity prototypes for real-world phenomena
  2. humans study the phenomena by interacting with the LLMs
- I am *very interested* in:
  - prompting strategies that elicit the relevant latent information in the models
  - querying methods that treat the model as a black box
- Bonus points if the work can be replicated in a classroom.
- I am *less interested* in:
  - NPCs and role-playing games, e.g., Tseng et al. [15]
  - white box and grey box methods, e.g., Schut et al. [12]

Embodied Agents

- Huang et al. [6] used LLMs to generate executable action plans for high-level household tasks without additional training.
- They evaluated the actions plans in the VirtualHome environment.[a]
- They showed a trade-off between executability and semantic correctness as judged by human evaluators.

[a] http://virtual-home.org/



LLMs can translate high-level tasks to medium-level steps.

The VirtualHome Environment

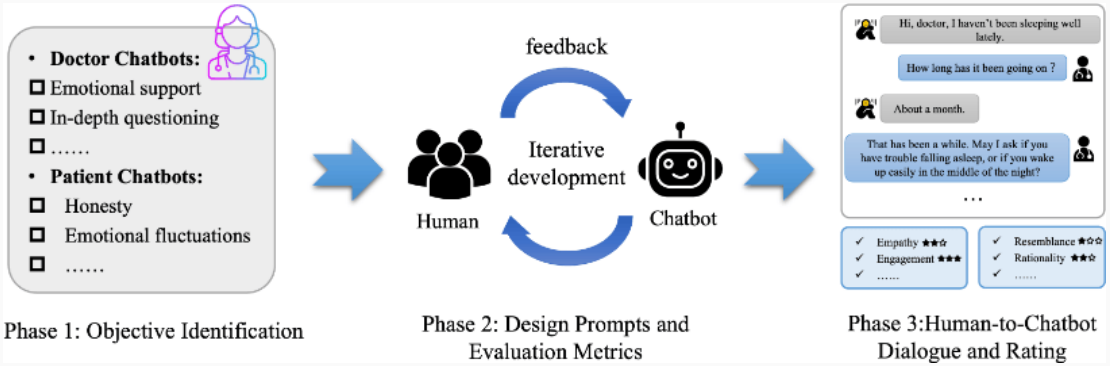
Social Simulacra

- Park et al. [8]: *Social simulacra* is a prototyping technique that relies on LLMs to generate realistic social interactions at scale in social computing systems (SimReddit).

- Park et al. [9]: *Generative agents* mimic human-like behaviour, such as waking up, cooking, socialising, and planning in a sand-boxed environment (Smallville).

Psychiatrist Conversations

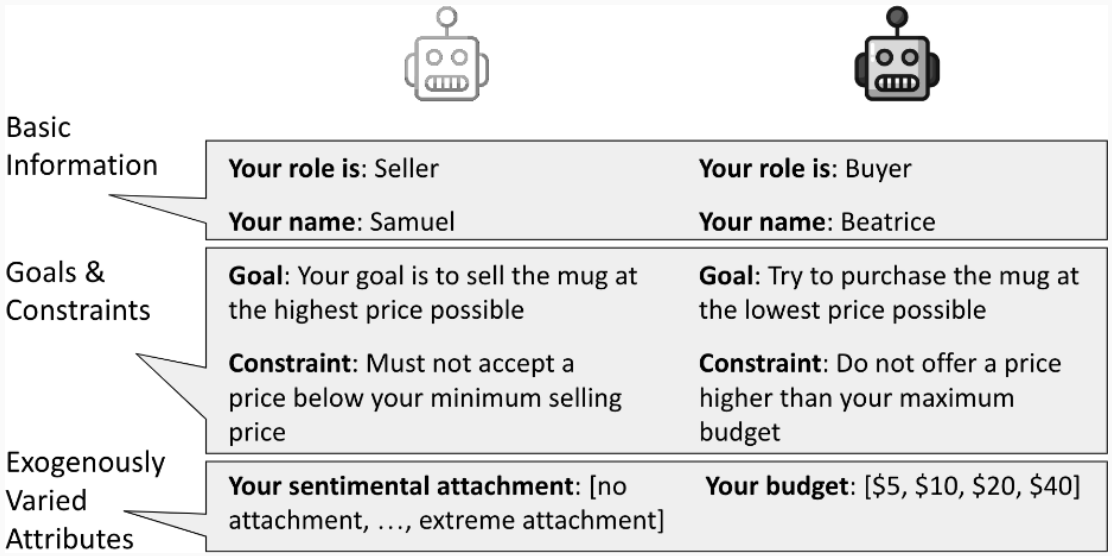
- Chen et al. [2] created *doctor chat-bots* that conducted diagnostic interviews with empathy and in-depth questioning, and *patient chat-bots* that exhibited emotional fluctuations, colloquial expression, and resistance to disclosure.



- Evaluation showed the importance of prompting strategies, e.g., excessive empathy leading to repetitive, template-like responses that degraded user experience.

Behavioural Economics

- Horton [5] investigates if LLMs can act as economic agents by replicating classic experiments (e.g., the dictator game) using LLMs with preferences and beliefs.



- Manning et al. [7] develop an automated system for generating and testing social science hypotheses.
- LLMs often ‘know’ more than they can articulate directly, but simulations can surface this latent knowledge.

References

[1] J. Chen, X. Wang, R. Xu, S. Yuan, Y. Zhang, W. Shi, J. Xie, S. Li, R. Yang, T. Zhu, A. Chen, N. Li, L. Chen, C. Hu, S. Wu, S. Ren, Z. Fu, and Y. Xiao. From persona to personalization: A survey on role-playing language agents.
https://arxiv.org/abs/2404.18231, 2023.

[2] S. Chen, M. Wu, K. Q. Zhu, K. Lan, Z. Zhang, and L. Cui. LLM-empowered chatbots for psychiatrist and patient simulation: Application and evaluation.
https://arxiv.org/abs/2305.13614, 2023.

[3] F. J. Dyson. Unfashionable pursuits. *The Mathematical Intelligencer*, 5:47–54, 1983.

[4] D. C. Engelbart. Augmenting human intellect: A conceptual framework.
https://www.dougengelbart.org/content/view/138 and
https://archive.ph/mYq5p , 1962.

[5] J. J. Horton. Large language models as simulated economic agents: What can we learn from homo silicus?
https://arxiv.org/abs/2301.07543 , 2023.

[6] W. Huang, P. Abbeel, D. Pathak, and I. Mordatch. Language models as zero-shot planners: Extracting actionable knowledge for embodied agents. https://arxiv.org/abs/2201.07207 , 2022.

[7] B. S. Manning, K. Zhu, and J. J. Horton. Automated social science: Language models as scientist and subjects.
https://arxiv.org/abs/2404.11794 , 2024.

[8] J. S. Park, L. Popowski, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein. Social simulacra: Creating populated prototypes for social computing systems. In *The Annual ACM Symposium on User Interface Software and Technology (UIST)*. ACM, 2022. doi: 10.1145/3526113.3545616.

[9] J. S. Park, J. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *The Annual ACM Symposium on User Interface Software and Technology (UIST)*. ACM, 2023. doi: 10.1145/3586183.3606763.

[10] V. Rao, T. Beiko, and J. Davis. *Protocol Reader: Summer of Protocols*. 2023. ISBN 9781962872805.

[11] V. Rao, T. Beiko, D. Ryan, J. Stark, T. V. Epps, and B. Aue. The unreasonable sufficiency of protocols.
↗ https://summerofprotocols.com/the-unreasonable-sufficiency-of-protocols-web and ↗ https://archive.ph/FNl13 , 2023.

[12] L. Schut, N. Tomašev, T. McGrath, D. Hassabis, U. Paquet, and B. Kim. Bridging the human–AI knowledge gap through concept discovery and transfer in AlphaZero. *The National Academy of Sciences of the United States of America (PNAS)*, 122(13), 2025. doi: 10.1073/pnas.2406675122.

[13] M. Shanahan, K. McDonell, and L. Reynolds. Role-play with large language models.
↗ https://arxiv.org/abs/2305.16367 , 2023.

[14] J. Stark. Atoms, institutions, blockchains. ↗ https://stark.mirror.xyz/n2UpRqwdf7yjuiPKVICPpGoUNeDhlWxGqjulrlpyYi0 and ↗ https://archive.ph/dzjIo , 2022.

[15] Y.-M. Tseng, Y.-C. Huang, T.-Y. Hsiao, W.-L. Chen, C.-W. Huang, Y. Meng, and Y.-N. Chen. Two tales of persona in LLMs: A survey of role-playing and personalization.
https://arxiv.org/abs/2406.01171, 2024.

Questions?