AllTrials.info

Making the clinical trial discovery process more accessible for patients through personalized trial matching using large language models

Dylan Fletcher, Conrad Chan, Shaki Pothini, Angela Ingram, Jonathan Hodges

Berkeley
UNIVERSITY OF CALIFORNIA

December 10, 2024

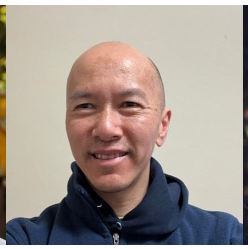
Image link:
https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.istockphoto.com%2Fphotos%2Ftwo-people-smiling-at-computer&psig=AOvVaw1U_QvG4LOQV87bT6PUs8LI&ust=1727206425565000&source=images&cd=vfe&opi=89978449&ved=0CBQQjRxqFwoTCLjS9cLn2YgDFQAAAAAdAAAAABAE

Note: The AllTrials logo was generated by DALL-E with ChatGPT.

Our Team

AllTrials


Dylan Fletcher


Conrad Chan


Shaki Pothini


Angela Ingram


Jonathan Hodges

Berkeley
UNIVERSITY OF CALIFORNIA

Solving Inequitable Access to Clinical Trials

AllTrials

- **Project:** AllTrials.info increases trial access.
- **Barriers:** Geographic locations of rural areas.
- **Solution:** Easier trial discovery for patients.
- **Impact:** Reduces costs, speeds access to life-saving treatments, potential increase in diversity.

Berkeley
UNIVERSITY OF CALIFORNIA

Note: image was generated by DALL-E with ChatGPT.

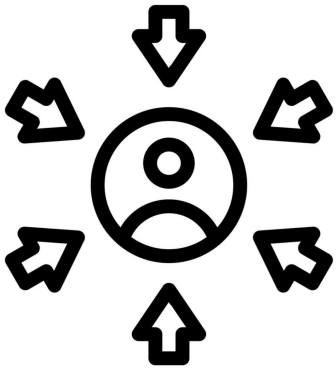
AllTrials.info is a differentiated solution

AllTrials

By utilizing large language models, we make it easier to find clinical trials that are right for patients.

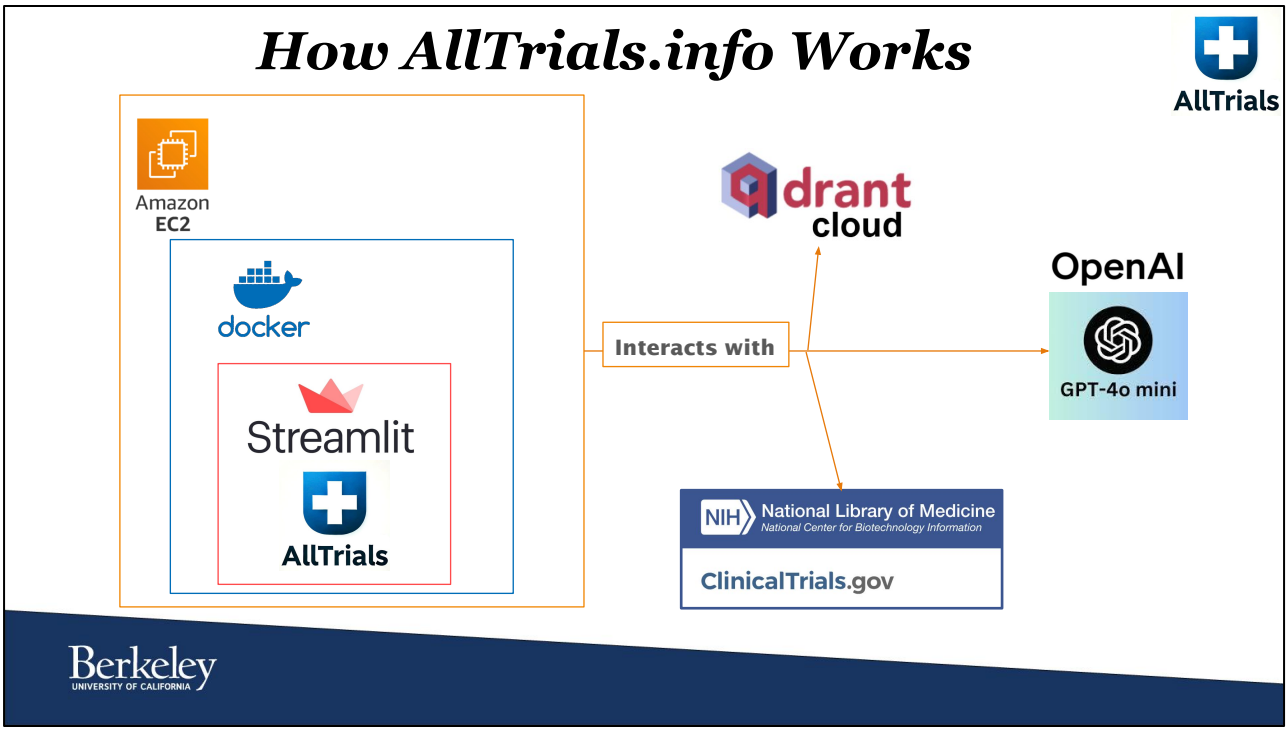
Link to AllTrials.info

Patient focused



Berkeley
UNIVERSITY OF CALIFORNIA

Icon link:
https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.vectorstock.com%2Froyalty-free-vector%2Fuser-centered-design-thick-line-icon-for-personal-vector-46444533&psig=AOvVaw3gTyCAoOqk7_4YhjRK1Ir4&ust=1730302831847000&source=images&cd=vfe&opi=89978449&ved=0CBcQjhxqFwoTCIix2cn2s4kDFQAAAAAdAAAAABAE



Docker image link:
https://www.google.com/url?sa=i&url=https%3A%2F%2Fmedium.com%2F%40anshulganvir%2Fintroduction-to-docker-337b9d09a079&psig=AOvVaw3Lsf_P-_IKRXSZJAdOf0R5&ust=1733340784638000&source=images&cd=vfe&opi=89978449&ved=0CBQQjRxqFwoTCMCQjeKrjIoDFQAAAAAdAAAAABAX

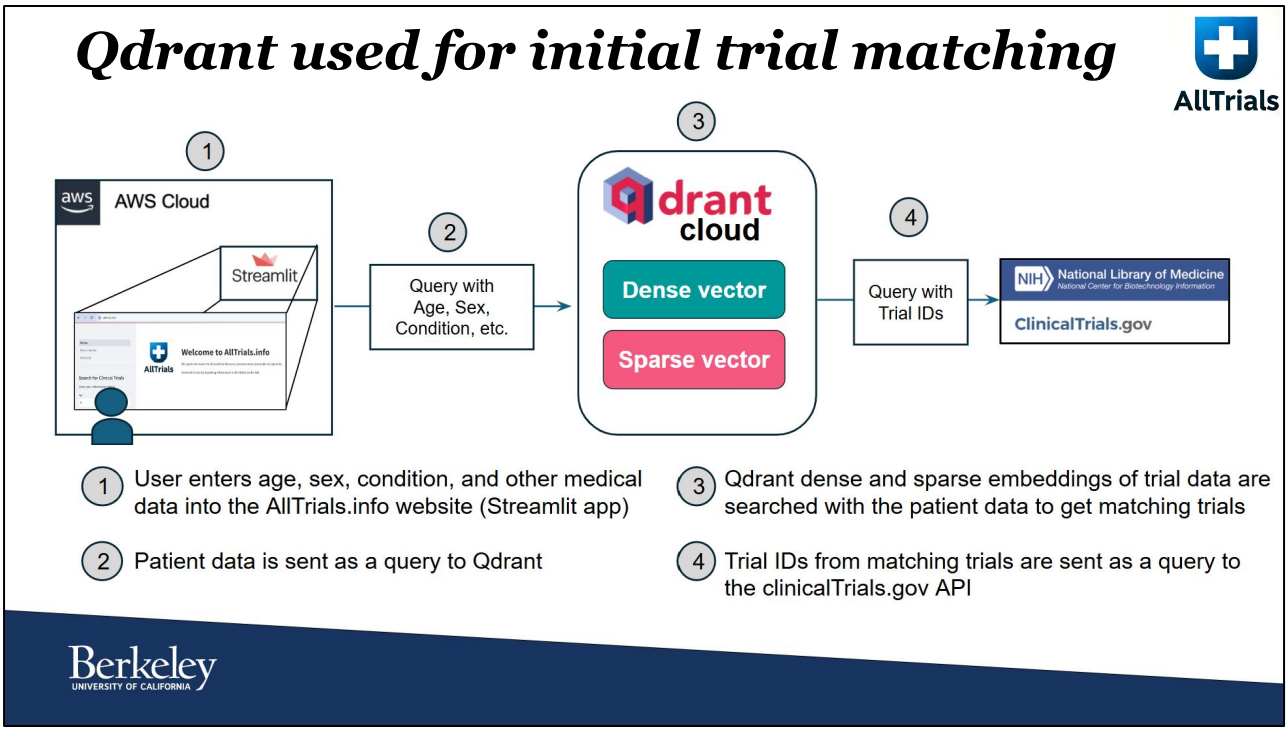
Amazon EC2 image link:
https://www.google.com/url?sa=i&url=https%3A%2F%2Fdev.to%2Faws-builders%2Famazon-ec2-everything-you-need-to-know-343j&psig=AOvVaw3YRJY1TPA9LHiRw2aHrlFh&ust=1733340886684000&source=images&cd=vfe&opi=89978449&ved=0CBQQjRxqFwoTCPCWzJCsjIoDFQAAAAAdAAAAABAE

Streamlit app image link:
https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.businesswire.com%2Fnews%2Fhome%2F20200616005364%2Fen%2FStreamlit-Raises-21M-in-Series-A-Funding-From-GGV-Capital-and-Gradient-Ventures-to-Amplify-the-Impact-of-Data-Science-and-Machine-Learning&psig=AOvVaw0IRcEe9gMsYw2D_elLTa3c&ust=1733340935460000&source=images&cd=vfe&opi=89978449&ved=0CBQQjRxqFwoTCNiW2qisjIoDFQAAAAAdAAAAABAE

ClinicalTrials image link: https://clinicaltrials.gov/

GPT-4o mini image link:
https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.hixx.ai%2Fblog%2Fxxai-news%2Fsurfacing-gpt-4o-mini&psig=AOvVaw2mC2Ajfz_HJaa1RsKH0moC&ust=1733417021000000&source=images&cd=vfe&opi=89978449&ved=0CBQQjRxqFwoTCNio-t_HjooDFQAAAAAdAAAAABAE

Qdrant image link:
https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.businesswire.com%2Fnews%2Fhome%2F20240416934769%2Fen%2FQdrant-Announces-an-Industry-First-Hybrid-Cloud-Offering-For-Enterprise-AI-Applications&psig=AOvVaw1FkUwSB9g2ctSAlwQFK41V&ust=1733417256764000&source=images&cd=vfe&opi=89978449&ved=0CBQQjRxqFwoTCMid2NLljooDFQAAAAAdAAAAABAE

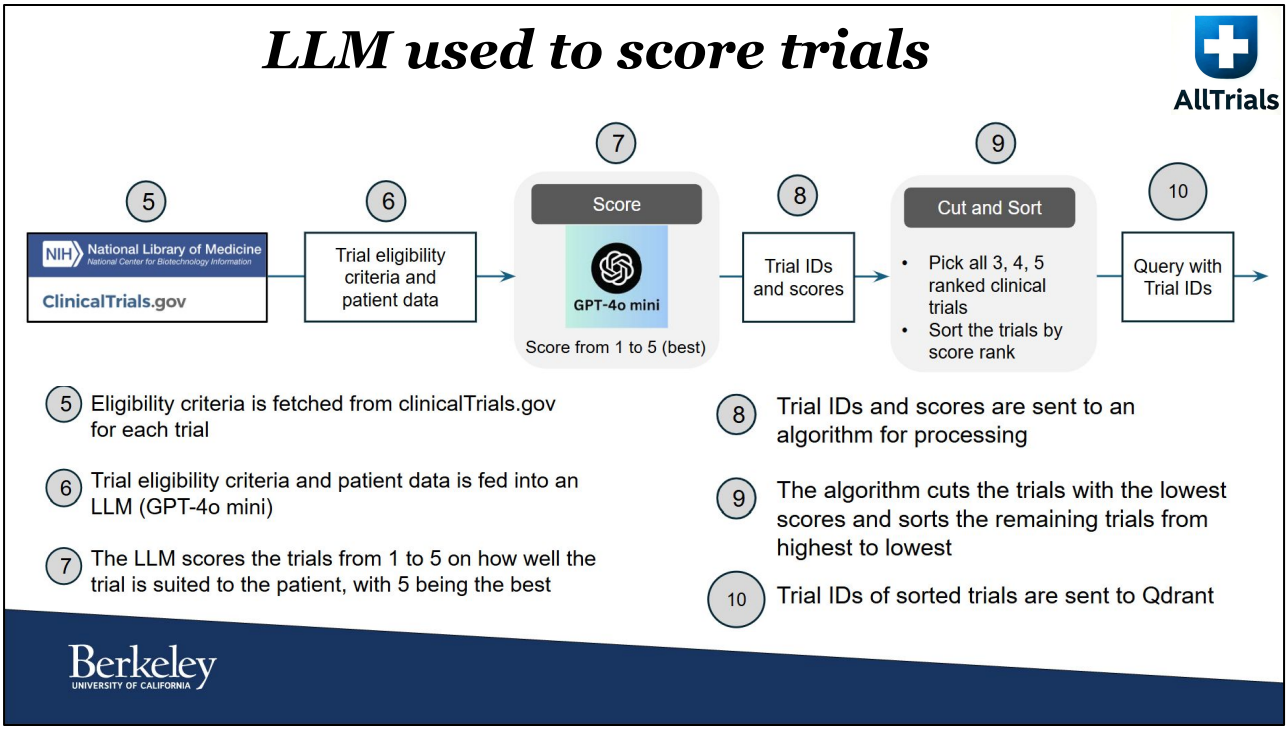


Streamlit app image link:
https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.businesswire.com%2Fnews%2Fhome%2F20200616005364%2Fen%2FStreamlit-Raises-21M-in-Series-A-Funding-From-GGV-Capital-and-Gradient-Ventures-to-Amplify-the-Impact-of-Data-Science-and-Machine-Learning&psig=AOvVaw0IRcEe9gMsYw2D_elLTa3c&ust=1733340935460000&source=images&cd=vfe&opi=89978449&ved=0CBQQjRxqFwoTCNiW2qisjIoDFQAAAAAdAAAAABAE

ClinicalTrials image link: https://clinicaltrials.gov/

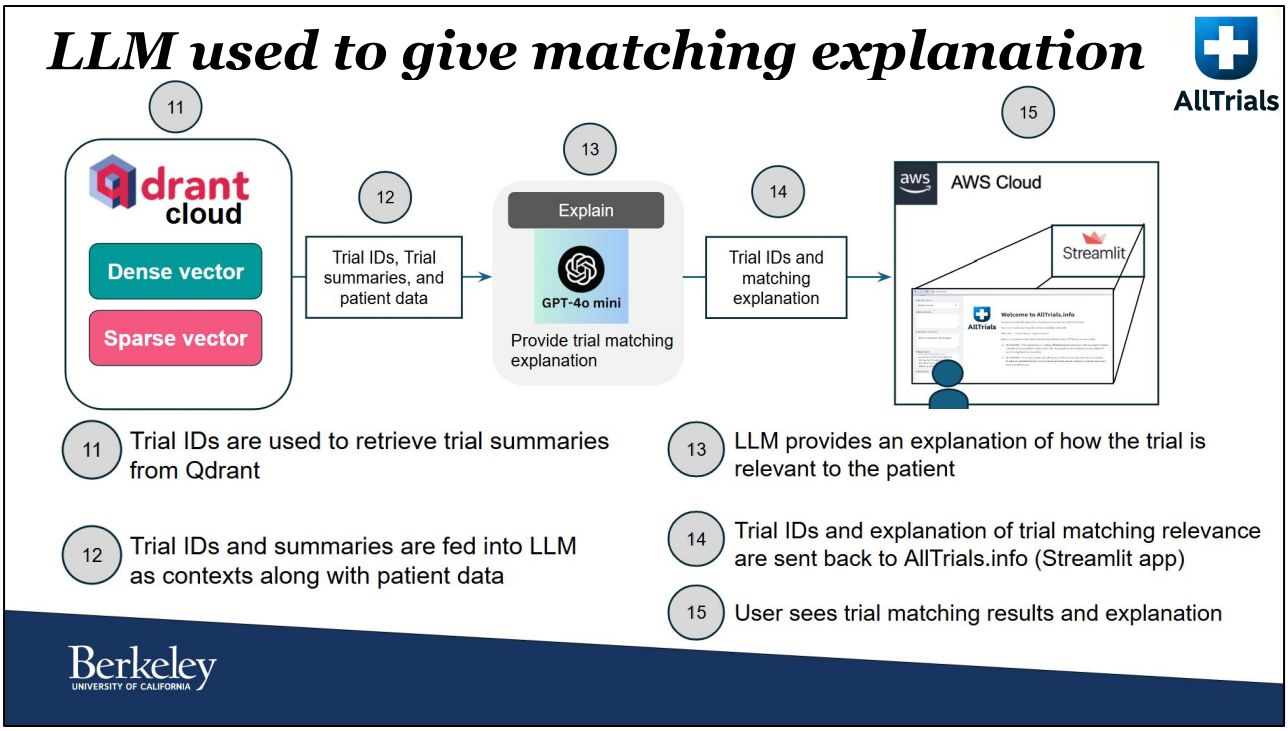
AWS Cloud image link:
https://www.google.com/url?sa=i&url=https%3A%2F%2Ftwitter.com%2Fjerryjliu0%2Fstatus%2F1744755577518256321&psig=AOvVaw1j0jO9YE_NB68w8R4KDI53&ust=1733417623940000&source=images&cd=vfe&opi=89978449&ved=0CBQQjRxqFwoTCKCMv4nKjooDFQAAAAAdAAAAABAb

Qdrant image link:
https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.businesswire.com%2Fnews%2Fhome%2F20240416934769%2Fen%2FQdrant-Announces-an-Industry-First-Hybrid-Cloud-Offering-For-Enterprise-AI-Applications&psig=AOvVaw1FkUwSB9g2ctSAlwQFK41V&ust=1733417256764000&source=images&cd=vfe&opi=89978449&ved=0CBQQjRxqFwoTCMid2NLljooDFQAAAAAdAAAAABAE



ClinicalTrials image link: https://clinicaltrials.gov/

GPT-4o mini image link: https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.hixx.ai%2Fblog%2Fxxai-news%2Fsurfacing-gpt-4o-mini&psig=AOvVaw2mC2Ajfz_HJaa1RsKH0moC&ust=1733417021000000&source=images&cd=vfe&opi=89978449&ved=0CBQQjRxqFwoTCNio-t_HjooDFQAAAAAdAAAAABAE



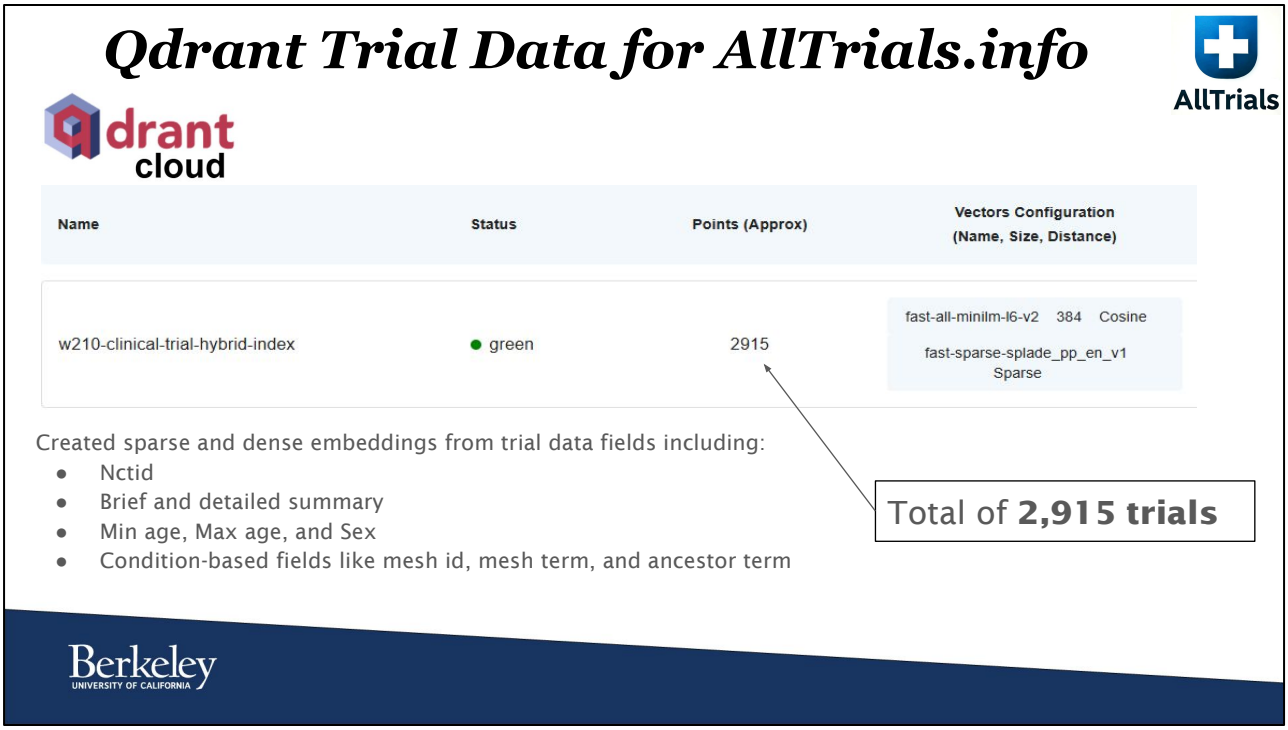
Streamlit app image link:
https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.businesswire.com%2Fnews%2Fhome%2F20200616005364%2Fen%2FStreamlit-Raises-21M-in-Series-A-Funding-From-GGV-Capital-and-Gradient-Ventures-to-Amplify-the-Impact-of-Data-Science-and-Machine-Learning&psig=AOvVaw0IRcEe9gMsYw2D_elLTa3c&ust=1733340935460000&source=images&cd=vfe&opi=89978449&ved=0CBQQjRxqFwoTCNiW2qisjIoDFQAAAAAdAAAAABAE

AWS Cloud image link:
https://www.google.com/url?sa=i&url=https%3A%2F%2Ftwitter.com%2Fjerryjliu0%2Fstatus%2F1744755577518256321&psig=AOvVaw1j0jO9YE_NB68w8R4KDI53&ust=1733417623940000&source=images&cd=vfe&opi=89978449&ved=0CBQQjRxqFwoTCKCMv4nKjooDFQAAAAAdAAAAABAb

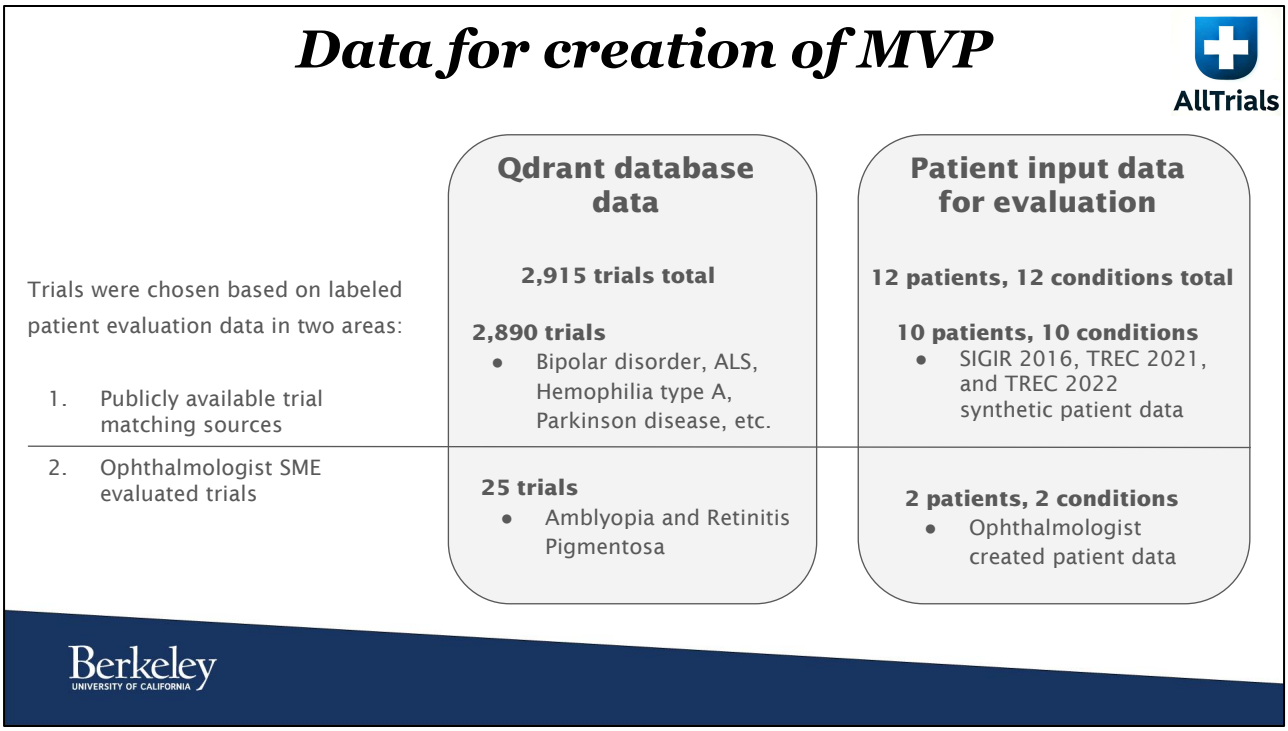
Qdrant image link:
https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.businesswire.com%2Fnews%2Fhome%2F20240416934769%2Fen%2FQdrant-Announces-an-Industry-First-Hybrid-Cloud-Offering-For-Enterprise-AI-Applications&psig=AOvVaw1FkUwSB9g2ctSAlwQFK41V&ust=1733417256764000&source=images&cd=vfe&opi=89978449&ved=0CBQQjRxqFwoTCMid2NLljooDFQAAAAAdAAAAABAE

GPT-4o mini image link:
https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.hixx.ai%2Fblog%2Fxxai-news%2Fsurfacing-gpt-4o-mini&psig=AOvVaw2mC2Ajfz_HJaa1RsKH0moC&ust=1733417021000000&source=images&cd=vfe&opi=89978449&ved=0CBQQjRxqFwoTCNio-t_HjooDFQAAAAAdAAAAABAE

Qdrant Trial Data for AllTrials.info



Qdrant image link:
https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.businesswire.com%2Fnews%2Fhome%2F20240416934769%2Fen%2FQdrant-Announces-an-Industry-First-Hybrid-Cloud-Offering-For-Enterprise-AI-Applications&psig=AOvVaw1FkUwSB9g2ctSAlwQFK41V&ust=1733417256764000&source=images&cd=vfe&opi=89978449&ved=0CBQQjRxqFwoTCMid2NLIjooDFQAAAAAdAAAAABAE



SIGIR 2016 data source: https://data.csiro.au/collection/csiro:17152

- Koopman, Bevan and Zuccon, Guido (2016). 'A test collection for matching patients to clinical trials' in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval,* p.669-672

TREC 2021 data source: https://www.trec-cds.org/2021.html

- Roberts, Kirk and Demner-Fushman, Dina and Voorhees, Ellen M and Bedrick, Steven and Hersh, Willian R (2021). 'Overview of the TREC 2021 clinical trials track' in *Proceedings of the Thirtieth Text REtrieval Conference (TREC 2021)*

TREC 2022 data source: https://www.trec-cds.org/2022.html

- Roberts, Kirk and Demner-Fushman, Dina and Voorhees, Ellen M and Bedrick, Steven and Hersh, Willian R (2022). 'Overview of the TREC 2022 clinical trials track' in *Proceedings of the Thirty-first Text REtrieval Conference (TREC 2022)*

Top Performing MVP Solution Results

AllTrials

For 10 synthetic patients using GPT-4o mini LLM

MODEL	Answer Correctness	Context Recall	Context Precision	Faithfulness	Noise Sensitivity
Qdrant Hybrid Search + LLM-based Eligibility Criteria Re-ranking	**0.9293**	**0.9420**	**0.9280**	**0.9333**	**0.176**

- **Answer Correctness** - compares and evaluates the factual accuracy of the generated response with the ground truth answer
- **Context Recall** - measures number of relevant chunks in context
- **Context Precision** - measures proportion of relevant chunks in context
- **Faithfulness** - measures the factual consistency of the generated response against the given context
- **Noise Sensitivity** - how often a system makes errors by providing incorrect responses when utilizing either relevant or irrelevant chunks in context

Berkeley
UNIVERSITY OF CALIFORNIA

RAGAS site for information on RAGAS metrics.

Final MVP Solution Journey

AllTrials

We compared different systems as candidates for the initial trial matching stage of our final solution.

	MODELS	Answer Correctness	Context Recall	Context Precision	Faithfulness	Noise Sensitivity
1	RAG - Static Neo4j Graph	0.8671	0.8631	0.8396	0.8118	0.2088
2	Qdrant Hybrid Search	**0.8843**	**0.9160**	**0.9080**	**0.8315**	**0.1868**
3	GraphRAG - Parquet	0.8458	0.4183	0.8057	0.3667	0.2433
4	GraphRAG - Neo4j	0.6943	0.3694	0.7813	0.3051	0.2882

Berkeley
UNIVERSITY OF CALIFORNIA

diollo Apothecary Scale
Amazon: https://m.media-amazon.com/images/I/71Qr-byNnAL._SX522_.jpg

Top 3 Technical Challenges

AllTrials

1. Re-ranking Trials Based on Free-Form Eligibility Criteria
2. Evaluation of Stochastic Systems with Natural Language Outputs
3. Scaling from Laptop to Full Application on AWS

Berkeley
UNIVERSITY OF CALIFORNIA

Brain Coach, Dr. Sunita Punjabi
San Antonio Woman Magazine:
https://www.instagram.com/sanantoniowoman/p/DBbhf3OJnO8/

Future Roadmap

AllTrials

- **Automate Trial Upload to Qdrant and Expand Trial Coverage**
  - Directly source active trials from ClinicalTrials.gov API
  - Include more trials and conditions
- **Improve Eligibility Scoring**
  - Parse individual criteria
  - Add user feedback option to site to improve model performance
- **Make UI more flexible**
  - Support user interactivity such as feedback inputs and navigation session storage
  - Investigate other platforms besides Streamlit for more customization

Berkeley
UNIVERSITY OF CALIFORNIA

Note: image was generated by DALL-E with ChatGPT.

Our Mission: Equitable Access to Clinical Trials for All

AllTrials

- **Mission:** Break down barriers to clinical trial access.
- **Impact:** Streamline access, accelerate treatments, foster diversity, and improve patient outcomes.

Berkeley
UNIVERSITY OF CALIFORNIA

Note: image was generated by DALL-E with ChatGPT.

AllTrials

Appendix

Berkeley
UNIVERSITY OF CALIFORNIA

Acknowledgements

We would like to thank our instructors, Korin Reid and Ramesh Sarukkai, for their guidance.

We would also like to thank Kevin Lee, M.D. for creating synthetic patients and identifying clinical trials for which those patients are eligible.