

# Research Scientist - NLP

Cambridge, MA / New York City, NY

Machine Learning – Machine Learning / Full-Time / Hybrid

## About Kensho:

Kensho is a 120-person Machine Learning (ML) and Natural Language Processing (NLP) company, centered around providing cutting-edge solutions to meet the challenges of some of the largest and most successful businesses and institutions. We are owned by S&P Global and operate independently. Our toolkit illuminates insights by helping the world better understand, process, and leverage messy data. Specifically, Kensho's solutions largely involve speech recognition (ASR), entity linking, structured document extraction, automated database linking, text classification, and more. We are continuously expanding our portfolio and are looking for passionate researchers to help us create state-of-the-art models across a variety of domains! Are you looking to solve hard problems and enjoy working with teammates with diverse perspectives? If so, we would love to help you excel here at Kensho. We are a collaborative group of experienced Research Scientists and Machine Learning Engineers, whose academic backgrounds include doctorate degrees in NLP, theoretical physics, statistics, etc. We take pride in our team-based, tightly-knit startup-like Kenshin community, which fosters continuous learning and a communicative environment.

At Kensho, we hire talented people and give them the freedom, support, and resources needed to accomplish our shared goals. We believe in flexibility-first and give our employees the opportunity to work from where they feel most productive and engaged (must be in the United States). We also value in-person collaboration, so there may be times when travel to one of our Kensho hubs (e.g., Cambridge, MA or NYC) will be required for team meetings or company events.

## About the R&D Lab:

Since 2022, we have been building a world-class R&D lab comprised of NLP Research Scientists, and we heavily prioritize publishing in top-tier conferences. Our small team has demonstrated compelling results

and is fueling innovation throughout Kensho and S&P Global at large. Specifically, we are continuously developing Large Language Models (LLMs) and are actively working on **long-context question-answering (QA)**, **complex reasoning**, **tokenization**, **alignment** (e.g., **factuality**), **multi-document QA**, and more!

**Our small team has reserved access to hundreds of fast GPUs (A100s), spanning Cloud and on-prem machines.**

Our current projects include:

- **Long-context document QA**, where the answer is contained within documents that are hundreds of pages in length [1]
- **Complex reasoning**, including better understanding and improving models' ability to approximate numbers (related to commonsense reasoning).
- Creating rigorous **evaluation benchmarks**, spanning domain knowledge, quantity extraction, and program synthesis [2]
- Improving existing **alignment techniques** for domain-specific needs, while also addressing factuality
- Dissecting **tokenizers** to better understand how each of the sub-components impact intrinsic and extrinsic performance [3][4]
- **Multi-Document QA** where the answer requires combining information from dozens of sources.
- Retrieval-augmented generation (**RAG**) methods
- Creating high-quality data filters for LLM development

Additionally, we maintain strong relationships with academia, including collaborating on several ongoing projects, providing industry grants, sponsoring conferences, and jointly holding faculty positions.

[1] [DocFinQA: A Long-Context Financial Reasoning Dataset](#) (Reddy et al., 2024)

[2] [BizBench: A quantitative reasoning benchmark for business and finance](#) (Koncel-Kedziorski et al., 2024)

[3] [Tokenization Is More Than Compression](#) (Schmidt et al., 2024)

[4] [Greed is All You Need: An Evaluation of Tokenizer Inference Methods](#) (Uzan et al., 2024)

Kensho states that the anticipated base salary range for the position is 150k-225k. In addition, this role is eligible for an annual incentive bonus and equity plans. At Kensho, it is not typical for an individual to be hired at or near the top of the range for their role and compensation decisions are dependent on the facts and circumstances of each case.

**Technologies & Tools We Use:**

- ML: PyTorch, Weights & Biases, NetworkX
- Deployment: Airflow, Docker, EC2, Kubernetes, AWS
- Datastores: Postgres, Elasticsearch, S3

## **What You'll Do:**

- Regularly reading late-breaking research papers and helping to identify pertinent directions of work
- Developing novel, state-of-the-art NLP models that can scale to millions of documents
- Working closely with other Research Scientists and ML Engineers
- Writing clean, readable research code in PyTorch (not expected to write production-level code)
- Contribute to a stellar engineering culture that values excellent design, documentation, testing, and code
- Share your research results with your colleagues (presentations) and the world (published papers, patents, and blog posts)

## **What We Look For:**

- Outstanding people come from all different backgrounds, and we're always interested in meeting talented people! Therefore, we do not require any particular credential or experience. If our work seems exciting to you, and you feel that you could excel in this position, we'd love to hear from you. That said, most of our successful candidates possess the following, which reflects both our technical needs and team culture:
- Hold a PhD in Computer Science or related field (or a Master's with significant research experience)
- Have published in a top-tier ML/NLP conference (e.g., ACL, NAACL, EMNLP, NeurIPS, ICML)
- Are proficient in writing code in PyTorch, Tensorflow, or JAX
- Have experience with the techniques required to work effectively with large, messy real-world data
- Prefer to collaborate iteratively on hard problems with your teammates rather than spending stretches of time working alone and presenting your results intermittently
- Have a love for learning new skills and domains
- Are excited to share knowledge freely, proactively, and effectively with others who are interested
- Are a generous teammate who takes work seriously without taking yourself too seriously

At Kensho, we pride ourselves on providing top-of-market benefits, including:

- Medical, Dental, and Vision insurance
- 100% company paid premiums
- Unlimited Paid Time Off
- 26 weeks of 100% paid Parental Leave (paternity and maternity)
- 401(k) plan with 6% employer matching
- Generous company matching on donations to non-profit charities
- Up to \$20,000 tuition assistance toward degree programs, plus up to \$4,000/year for ongoing professional education such as industry conferences
- Plentiful snacks, drinks, and regularly catered lunches
- Dog-friendly office (CAM office)
- Bike sharing program memberships
- Compassion leave and elder care leave
- Mentoring and additional learning opportunities
- Opportunity to expand professional network and participate in conferences and events


## About Kensho

Kensho is an Artificial Intelligence company that builds solutions to uncover insights in messy and unstructured data that enable critical workflows and empower businesses to make decisions with conviction. We were founded in 2013 and now serve as S&P Global's innovation hub. We continue to maintain our distinct, independent brand in order to best promote our breakthrough, innovative culture. Our team of Kenshins enjoy a dynamic and collaborative work environment that runs autonomously from S&P Global, while leveraging the unparalleled breadth and depth of data and resources available as part of S&P Global. As Kenshins, we pride ourselves on maintaining an innovative culture that depends on diversity and inclusion.

We are an equal opportunity employer that welcomes future Kenshins with all experiences and perspectives. Kensho is headquartered in Cambridge, MA, with an office in New York City. All qualified

applicants will receive consideration for employment without regard to race, color, religion, sex, sexual orientation, gender identity, or national origin.

[Kensho Home Page](#)

Jobs powered by  **LEVER**