2021 ML Team Descriptions NYC & BOS

Facebook NYC

EXPERT VOICES

The new Expert Voices team focuses on helping independent journalists and experts to be successful on Facebook – supporting their distribution and also building a platform to help them create content and monetize. It's a new team just getting off the ground, and it's full-stack.

Ads Core ML

Ads Core ML team's mission is to maximize the value of advertising delivered to advertisers and Facebook users, through building scalable machine learning systems, innovating cutting-edge machine learning algorithms, and formulating machine learning optimization problems to enable next-generation ads products. The expertise of our engineering team covers disciplinary research and engineering areas including distributed systems, deep learning, embedding algorithms, recommender systems, multi-variable optimization, explore / exploit, user modeling and intent understanding, image and text content modeling and understanding, sparse modeling, auto ML, hardware accelerators for training and inference. Our team is responsible for creating long-term value for business and users via building both core technology innovations and enabling new products.

The team is working across the ranking stack and is making contributions on both ML and infrastructure sides, making the overall system more stable and better optimized for ads ranking. Please consider joining Ads Core ML, if you are interested in working on cutting edge ML techniques on a scalable system that serves billions of users every day.

PLACES DATA

Every month, over 1 billion people search, learn, engage and recommend places on Facebook and Instagram. To give them the best experience, our team leverages various Machine Learning techniques, and large scale algorithms to learn the best representation of public places around the world.

We manage and organize a massive graph of hundreds of millions of location nodes and billions of different edges connecting between them. We design and execute experiments to test the effectiveness of our work. We measure our success by leveraging both online and offline metrics to indicate the performance of our models and their effect on user engagement.

Technologies: Classification, Clustering, Data Mining, Python, Hack (PHP), Dataswarm, Hive, Presto, Spark, FBLearner, Fluent2

PLACE LOCATION CONTEXT (PLC)

The Personal Location Context (PLC) team builds privacy-aware machine learning systems that produce meaningful inferences about a person's *location context*: Are they at home? Are they on their way to work? Are they walking or driving? Did they just arrive at a restaurant? Are they shopping? What coffee shops have they visited in the past? Which is their regular grocery store?

We use a variety of sensor data (accelerometers, ambient WiFi signals, GPS locations, gyroscopes, and even video and audio) on a broad range of platforms (from mobile phones to future AR devices).

All of our work considers privacy first, and happens within unique computational constraints: How do we build a system that provides useful inferences while collecting as little data as possible? How can we infer a person's context quickly, constantly, and accurately with minimal battery usage? How can we leverage large-scale data while ensuring that sensitive information never leaves a person's device? How do we design a maximally reusable system that runs across very different platforms?

To tackle those challenges, we are working with new privacy-preserving machine learning techniques; moving computation and storage on-device; exploring novel methods for data minimization, deletion, and encryption; and much more.

Ultimately, our inferences aim to enrich people's interactions with the real world and make their experiences more meaningful while protecting their privacy. Those experiences are still being imagined, and our team is defining what the future of privacy-first, location-based augmented reality will look like.

Technologies: Classification, Statistical modeling, Outlier detection, Generative models, Calibration, Python, C++, Hack (PHP), Dataswarm, FBLearner Flow

KNOWLEDGE MANAGEMENT

The Knowledge Management team's goal is to organize workplace information and make it more efficiently accessible. We have ML systems for classification of information and mapping it to taxonomies. The team is also heavily invested in information retrieval and ranking systems that solve enterprise search. Knowledge recommendation is another area of focus of the team where we are building recommendation systems to suggest relevant information at the right place and time to our users. Across these systems, the team leverages the following technologies:

Technologies: NLP, Information Retrieval, Recommendation Systems, Deep learning, Clustering, Classification.

NEWSFEED AND STORIES RANKING

The Feed and Stories Recommendations team provides foundational ML capabilities for many of the most popular recommendation apps across Facebook, including News Feed Recommendations. We work on content understanding to improve the ranking and retrieval of recommendations, and develop techniques to help identify and promote valuable creators & their content.

Technologies: Recommender systems, NLP, personalization, ranking, retrieval, content understanding, user understanding

References (Public Links):

- https://code.facebook.com/posts/1072626246134461/introducing-fblearner-flow-facebook-s-ai-backbone/
- https://research.fb.com/category/facebook-ai-research-fair/
- https://research.fb.com/downloads/starspace/
- https://research.fb.com/downloads/fasttext/

NEWS ECOSYSTEM AND RELEVANCE (NEAR)

The News Foundations team is building Machine Learning models, classifiers, ranking interventions and infrastructure to address the distribution of news across Facebook. We do this by building deep content and user representations and using these representations to identify news content that is most relevant to individuals, and is aligned with internally defined News Principles. On a typical day, engineers on our team might be building a new model/classifier, creating state of the art news document representations, running/analyzing ranking experiments in Facebook's news feed or building performant infrastructure to serve our algorithms.

NEWS TAB RECOMMENDATIONS

Facebook News is a dedicated place for news on Facebook that is currently available in the US, UK, and launching in more countries later this year.. The current news experience in Feed is unpredictable and the

credibility of the sources shown can be questionable. These are the reasons we are building a dedicated space for news.

The News Tab Recommendations team powers the ranking of personalized News recommendations on the News Tab. The team is composed of three pillars, focused on Infra (perf, API & real-time execution), Relevance (retrieval, ML, ranking), and Quality (objective quality signals).

One-pager: https://fb.quip.com/hW5IACId5WQq
Eng wiki (WIP): https://fburl.com/wiki/574bihzf

DATING RANKING

Facebook Dating is a quickly growing online dating product. Our mission is to make it the best place for people to find & build meaningful relationships. As such, ranking is the most critical piece to the equation as it is the recommendation engine to suggest matches.

We are only at the beginning of our journey as this ML space is very unique and unexplored. We are looking for strong ML & Ranking EMs who are not afraid to go technically deep, apply state-of-the-art research, and are driven to innovate.

Civic Integrity

Civic Integrity brings some of the toughest challenges and questions on and off social media platforms. We as a team, have been dedicated to the mission of giving people voice and power to build a civically-engaged community with our integrity efforts. We are building end-to-end solutions to understand, detect and mitigate different types of integrity problems related to civic and elections. ML techniques are broadly applied through our work to help us achieve these goals.

In New York, we are focusing on designing and building several models to solve problems such as civic entity understanding, integrity issues detection and ranking intervention. We believe ML can take some of our existing solutions to new levels of scalability, efficiency and accuracy. We are looking forward to working with experienced ML engineers who are excited to drive such efforts and help lead us to the next level of success.

INFERENCE PLATFORM

Our team's mission is to provide a reliable and efficient managed hosting environment to run Facebook's state of the art machine learning models. We build cutting edge infrastructure and distributed systems to run thousands of models for hundreds of different teams and use cases across Facebook. The team is made up of backend engineers and full stack engineers working together to build a platform that makes the process of deploying and using ML models in production as simple as an API call. We are working on many interesting distributed systems problems at the heart of the machine learning space at Facebook.

RAI TRANSPARENCY & CONTROL

Our team's mission is to help the Facebook family of apps provide the right level of transparency and control to the right stakeholders, internal and external, for all AI Experiences. The team builds tools, guidelines and frameworks that create a set of best practices and works to gain adoption across Facebook. We also publish some of our work externally to help the external Responsible AI community. Our mission is complete when all users of our apps trust Facebook AI, have confidence in the level of control provided to them, and understand how their actions affect the automated decisions making in our systems.

Our team consists of Research Scientists, Product Engineers and ML Engineers. We are very collaborative and we work with many different teams at Facebook and we have a strong cross functional team of Data Scientists, Data Engineers, Designers, User Researchers. Here is a sample project from our team:

Captum.ai - A PyTorch library which supports interpretability of models across modalities. Our team supports this library by adding new interpretability and explainability methods based on our own and external research. Our team also does active research to help advance XAI and develop new techniques for internal and external ML engineering teams.

References: https://captum.ai

Instagram

INSTAGRAM RECOMMENDATION FOUNDATION ENGINE

The Recommendation Relevance Foundation team is responsible for improving foundational technologies to improve unconnected recommendations that power Reels and Explore. We do so by both, working on end-end projects to help drive impact, and creating key central ML artifacts that can unblock projects across multiple work streams in IG. Our current focus spans two broad areas: Content Understanding and Delivery.

For Content Understanding, we are responsible for ensuring our ranking pipelines leverage state-of-the-art content understanding signals. We do so by partnering with research teams and improving and building models where necessary. Some of our key initiatives in 2021 H1 include topic and emotion modeling, defining and promoting genuine reels and working on online media clustering features.

For Delivery, our goal is to ensure that our multi-stage recommendation pipeline is responsive, reliable, and efficient. In 2020 H1, we are moving towards this goal by working on Fast Personalization, reconsideration ranking by offline caching of ranked results, Recommendation Metrics, and measurement and improvement of end-end latency

INSTAGRAM DISCOVERY INTEGRITY

Discovery Integrity is a surface integrity team focused on “intentful” products at Instagram e.g. Search, Hashtags, Locations and more. Discovery surfaces are among the most visited on Instagram, and face unique integrity challenges such as:

- How can we hold a high bar for results quality, while not hurting relevance and utility?
- How can we connect people with safe and high quality content around niche topics, without limiting community voice?
- When people are actively seeking risky results, how can we guide them towards less risky results and connect them with authoritative information?

We’re looking for ML engineers passionate about working on a highly collaborative cross-functional team to ensure that intentful discovery surfaces across Instagram connect users with delightful and safe experiences.

INSTAGRAM SEARCH RANKING

Our team's mission is to help you connect with your friends and family, interests, and local businesses on Instagram. We build scalable systems with cutting-edge machine learning to help hundreds of millions of users per day find what they're looking for. We are a full-stack team - we work on everything from designing the best user experiences for interest discovery, to building personalized ranking and recommendations models. Here's a sample project from our team:

- Personalized search ranking - There are millions of people named John on Instagram, but when you search for “john”, we want to show you the 5 people you're most likely to be searching for. Our search infrastructure leverages the social graph to do personalized retrieval, which is further personalized via machine learned ranking models, to bring you the most relevant results nearly instantly.

Instagram Search Ranking is a small team that powers a massively used product: IG search is crucial for helping people connect with people and hashtags on Instagram. Our long-term strategy is to help you find anything anywhere on Instagram. For the first time ever, we're building a way for you to search content via open-ended keyword search - photos, videos, stories, and more. Whether it's helping sneakerheads find the best kicks or connecting nail art fans with awesome nail art tutorials, search will deliver the most engaging and relevant content for any intent on Instagram. We have a lot of exciting, complex infrastructure and

ranking problems to work through. Engineers on our team focus on product-oriented ML work: building training pipelines, defining online metrics, creating offline human rater evaluations, training and tuning ranking models, iterating rapidly via online experiments, and building new search features.

Technologies: relevancy, ranking, personalization, retrieval

INSTAGRAM CONTENT INFRA TEAM

IG content Infra team is responsible for maintaining an efficient, reliable, scalable, low latency platform that supports real-time content retrieval and ranking via MezQL (a query language we built for content retrieval) under massive QPS.

Now that we've successfully onboarded the majority of IG products, we are expanding our scope to combine the best from FB and IG infrastructure, moving to a shared and more interoperable architecture to support launching products across FB & IG family via MezQL.

Furthermore, we are leveraging advanced inference technologies from FB to empower the platform with more capacities to execute modern ML tasks.

The team focuses broadly on workstreams revolving around efficiency, reliability, improved developer experience, and consolidation of infrastructure. The team's primary stack are C++, Python and Lua,. We have positions across levels and an excellent mentorship base to ramp you up quickly, so reach out and let's chat!

INSTAGRAM EXPLORE

Instagram Explore is one of the world's biggest social discovery platforms – more than half of accounts on Instagram use Explore every month. Our goal is to recommend people the best content across Instagram that fit their interests. More than 50% of accounts on Instagram utilize Explore every month!

The Instagram Explore Ranking team works on a wide variety of challenging ML problems including personalization, ranking, content retrieval, content understanding and user interests. We also closely work with other ML teams such as IGML Infra team on challenging infra problems, e.g., how to efficiently rank content from billions of posts, and how to maintain system reliability with extremely large traffic to our recommendation service.

Technologies: personalization, deep learning, embedding, recommendation systems

INSTAGRAM FEED RANKING

Our team is responsible for the Ranking Instagram Feed content. We are right at the center of Instagram and the decisions that we make impact every Instagram user. This gives us significant leverage when it comes to driving our mission of improving user experience by showing people the content that matters to them most.

INSTAGRAM STORIES RANKING

IG Stories Ranking team builds ranking algorithms that help bring most relevant and most engaging Stories content in front of our users. We use Machine Learning to understand what matters most to our users and build solutions that improve their experience with Stories - one of the most used and top revenue generating surfaces at IG.

Some of the directions that we are engaged in include incorporating new signals into our models, improving personalization, using new models and algorithms to improve ranking, enabling content creation though ranking.

Technologies: ML, ranking, personalization

INSTAGRAM PERSONALIZATION TECHNOLOGY ENGINEERING

We focus on making machine learning and personalization techniques widely available to Instagram products. This includes working on model architecture and their related infrastructure, offline evaluation and optimization and finally some specific modeling technique for user understanding.

Modeling: Deep Learning, Sequence Modeling, User Understanding
Stack: C2 and Pytorch

INTERESTS CONTENT QUALITY

The Interests Content Quality (ICQ) team is focused on ensuring we are recommending high quality unconnected content on Instagram Explore and Hashtag pages. The team uses machine learning to classify content as low quality and depending on confidence either down rank or filter it. The team is focused on developing both media level and account level signals to help in this process. These signals come from a variety of sources. The team uses point wise media level classifiers. They also develop embeddings to cluster non-recommendable accounts and hashtags.

Technologies: relevancy, classification, ranking, personalization, filtering

INSTAGRAM PRODUCER VALUE ENGINEERING

Our mission is to make Reels the leading platform for short-form entertaining video. The Reels Producer Relevance team is focused on inspiring and rewarding Reels creators to ensure that our content marketplace is full of rich, entertaining and diverse content. Our strategy is to incentivize sharing by enabling everyone to find their audience and get positive feedback, and giving them a chance to break when they make something great.

We are looking for ML Engineers who are passionate about working on recommendation systems and solving the problem on the producer side, which has significantly more challenges than consumption side problems.

INSTAGRAM PRODUCT NOTIFICATIONS TEAM

Notification Systems Mission Instagram Notification Systems (IGNS) team exists to create and maintain the most accessible platforms delivering the value of Instagram. Charter IGNS is responsible and accountable for, creation, support and enhancements of the platforms used to facilitate communication between Instagram users, Instagram teams to Instagram users, as well as automatically triggered communications, such as post likes, comment updates etc. Our platforms support multiple communication channels including in-app, push, SMS and email, and cover on-demand as well as recurring communications (e.g. notification campaigns).

Our objectives are:

- Create centralized easy-to-use, reliable and scalable systems enabling end-to-end communications life cycle management across all channels without any development effort.
- Create user-facing communications experiences aligned with Instagram principles.
- Establish quantitative and qualitative metrics for communications within Instagram and use these metrics to promote high quality user experience.

INSTAGRAM FEED RECOMMENDATION RELEVANCE ENGINE

Our team recommends new content (aka Feed Recs aka Suggested Posts) to our 1 billion+ daily Instagram Feed users via a blend of retrieval, ranking, content understanding and distributed processing. The team operates at true internet scale to find the best, personalized content for each user.

INSTAGRAM RECOMMENDATION CORE RANKING

The IG Recommendation Core ML team builds the machine learning models that power the ranking of IG's recommendation systems, primarily focusing on IG Reels, but also IG Explore. We are a team of Research Scientists and ML engineers. We aim to power Reels with the most relevant recommendations to

- Provide a great entertaining experience to users and
- To make the best videos thrive and reach large audiences, regardless of the creator's popularity.

In order to improve the recommendation relevance, we use cutting-edge recommendation ML techniques, such as large scale neural network modeling, multi-stage ranking, value modeling, transfer learning, user history modeling, etc.

INSTAGRAM SHOPPING TAB RELEVANCE

Team Mission Instagram Shopping is an $1.3 trillion opportunity (Facebook Shares Reach All-Time High). Our team mission is to build the world's most entertaining and engaging shopping experience with advanced personalization technologies. We own the end-to-end recommender system that powers Instagram Shop Tab including retrieval (sourcing), filtering, ranking and its delivery infra, across multiple shopping formats, such as products, brands, media, videos, collections, guides, live, IGTV, reels, drops etc.

We are looking a ML Ranking Engineers, who are passionate about shopping on Instagram and our team mission is looking for challenging problem space with high impact and visibility of leading a company-level top priority project wants an opportunity to lead a small team of engineers to design and deliver the most complex shopping ranking system in FB and IG (bonus) has prior experience in building e-commerce recommender system at scale (bonus) is interested in injecting your own product instinct to influence product strategy

Technologies: personalization, recommendation systems

INSTAGRAM SHOPPING SEARCH

The shopping search team is new as of 2020. We are building out intentional discovery surfaces for people to connect to the brands and products they love on Instagram.

Are you excited about building 0→1 experiences? Do you wish you could search for the trendiest new kicks to buy on Instagram? Are you passionate about ML, infra or product and want to work on Search? On Shopping Search, we have a unique opportunity to build a new kind of visual shopping search engine powered by brands, creators, and the data from millions of Instagram users. In this team you will build state-of-the-art solutions to help establish Shopping as a business for Instagram. In particular, you will work on a variety of technologies for text and visual understanding, retrieval, ranking and product for projects like new IG Shop Tab, Recommendation Systems, Visual Search. Beyond this, there are plenty of opportunities to take on strategic projects that cut across Instagram Shopping. We work closely with teams around the world, including FB Marketplace, Commerce AI, Catalog Platform, Instagram Shopping and more

Technologies: relevancy, ranking, personalization, retrieval

INSTAGRAM ML

The team builds a comprehensive platform that powers all of Instagram's ML development and production needs across Ranking, Sourcing, Recommendations and data. For Instagram's ML developers, this is a complete ecosystem that allows them to efficiently focus on the ML problem space. The platform provides tooling and frameworks to simplify and speed up Instagram specific ML workflows and experimentation. This platform is a top notch production environment that makes real-time predictions under massive QPS and flawlessly integrates ML into their experience.

Our primary stack is C++. ML knowledge is nice to have but not required for most roles.

Facebook Boston

LOCAL RELEVANCE

The Local Relevance team's mission is to build an end-to-end ML platform and portfolio of location features used by product teams in Facebook Reality Labs and across the Facebook Family of Apps to identify, retrieve, and rank locally relevant content. Our team works closely with teams in Spark AR on shipping 0=>1 experiences that will shape the future of augmented reality. We are a hybrid infrastructure / ML team that owns both the feature engineering of location features and platform that makes it possible. We also own the platform that stores, queries, and ranks world-anchored AR content.

Technologies: relevancy, ranking, personalization, retrieval, python, sql, c++, hack

WAREHOUSE DATA PRIVACY

The team is responsible for building systems and frameworks in our Big Data infrastructure to enable Facebook to provide enforcement to honor internal data policies, regulatory commitments such as GDPR and CCPA, and most importantly, our commitment to our end users that we responsibly store and use their data.

HOME LOCATION

The Home Location team's mission is to build the world's best infrastructure to understand user location. The team builds unified Location APIs for the Facebook family of apps that puts user location information in one place and provides rich location features and inferences (e.g. travel prediction, home prediction) that will continue to be enhanced over time. In addition, Home Location builds models to understand if a user is faking their location, which is important in preventing spam, election interference and other bad actions on the platform.

Home location uses machine learning to create predictions that power and improve end-user experiences in products such as Facebook Marketplace, Instagram search and more. This person will work with a very strong hybrid infrastructure/ML team and help define the ML roadmap for Home Location. They will also work closely with NYC, Menlo Park, and Seattle ML engineers.

Technologies: GBDT, CNN, RNN, HMM, mixed / ensemble, Hack, FbLearner Flow, Python

SPATIAL COMPUTING

The World.AI team improves the global map within Facebook with ML data extracted from other sources. We use the satellite imagery and deep neural segmentation nets to locate the missing or incorrect roads and buildings. We collaborate with the open-source mapping community, academia, NGO’s and industry to create the most detailed and relevant maps and spatial datasets.Some applications:

- **Global Population Density**: Using a mixture of machine learning techniques on high-resolution satellite imagery, the World.AI team identifies buildings and combines the results with existing census counts from the Center for International Earth Science Information Network (CIESIN) at Columbia University. The result – the world's highest-resolution population density maps, published openly in order to improve how nonprofits and NGOs do their work, how researchers learn, and how policies and infrastructure are developed.
- **Improving OpenStreetMap**: We use satellite imagery and deep neural nets to add roads and buildings to Open Street Map. Conversely, we can use it as a literal ground truth to make sure that when we incorporate Open Street Map edits into Facebook's map, we know those edits correspond to

real features on the ground. We collaborate with the open-source mapping community, academia, NGO's and industry in order to create the most detailed and relevant maps and spatial datasets.
- **Pushing forward the state of the art:** We engage with researchers from the computer vision community to push forward the state of the art in learning from satellite imagery.

Public Link: https://mapwith.ai

Technologies: Classification, Segmentation DNNs, Computer Vision, Graph theory, PyTorch, OpenCV, FBLearner Flow

Facebook AI Research (FAIR)

Facebook AI Research (FAIR) is dedicated to advancing the state of the art in machine intelligence through open research. We are primarily focused on fundamental research in a range of areas including Computer Vision, Natural Language Understanding, Dialog Systems, Reinforcement Learning, Machine Translation, and Speech.

RESEARCH ENGINEERING

Our Research Engineering team works in concert with FAIR Research Scientists to accelerate the research process, either through working directly on the research, or by building tools and technologies such as higher-level frameworks on top of PyTorch that make the research process faster and more productive.

Technologies**:** Deep Learning, DL Frameworks

AI APPLIED RESEARCH RELEVANCE

The goal of the Applied Research Personalization team is to build cutting-edge AI technologies to connect people to the content, information, and communities they care about. To achieve that goal, we push the state of the art in applied AI/ML research to learn from massive and diverse datasets in a transparent, controllable and privacy-aware manner, and build powerful and flexible frameworks that helps us train and deploy the best models at Facebook scale. Our algorithms and frameworks are used by hundreds of engineers across Facebook, Instagram, and Messenger and have a real impact on billions of people by powering News Feed, Search, Ads ranking, and other products and services across the Facebook ecosystem.

Specifically, the team in NYC is focused on designing and building next-gen models and tools to power safe, secure and highly personalized user-centric experiences. Sounds exciting? We'd love to hear from experienced ML researchers and engineers who want to help us achieve these big goals.

Technologies: Deep Learning, Representation Learning, Graph Theory, Sequence Modeling, Caffe2, PyTorch

AI PLATFORM - ML PIPELINE

The mission of ML Pipelines in Facebook AI is to deliver an easy-to-use, scalable, reliable, state of the art E2E ML platform that enables new AI research breakthroughs, accelerating quick iteration from idea to production. We are the TFX of Facebook; creating a unified ML programming model for reproducible pipelines with an expressive DSL, an IR (intermediate representation) to enable multiple execution environments (Batch vs realtime vs Interactive), and a standard set of libraries covering ML lifecycle from data sources, features to training and publishing with AutoML capabilities. Our customers are the developers in business critical teams across Facebook who want to prototype and productionize their ML models in the FB ecosystem.

AI PLATFORM - PYTORCH

PyTorch Domains team's mission is to accelerate research through novel, production-ready building blocks.

1. Accelerate external research: Provide reusable, orthogonal, correct, and performant building blocks for cutting-edge experimentation based on deep knowledge of the research domains and communities. Engage with the communities by supporting the development of novel models and best practices.

2. Transfer to production: Deeply integrate with the broad range of PyTorch capabilities, such as jit, quantization, distributed, and mobile, to enable seamless research-to-production for core end-to-end applications. Collaborate with domain-specific teams in Applied Research, FAIR and select external partners.
3. Discover novel 10x adoption opportunities: Identify and evaluate novel techniques and toolsets by developing deep technical understanding within our team and building new research collaborations.

Here are some of the domain libraries:
- https://github.com/pytorch/audio
- https://github.com/pytorch/vision
- https://github.com/pytorch/text

FAIR INFRA

Facebook AI Research Infra (FAIR Infra) team works on building cutting-edge AI Research Infrastructure. Every term in the name FAIR presents a unique set of challenges on Infra. AI workloads present requirements that are not ideal for traditional infra (GPU+low latency network, batch processing, custom storage). Research landscape is constantly changing, needs infra that is agile and supports fast & easy prototyping . Lastly, at Facebook we endeavor to take proven research ideas to our billions of users quickly.

All of these requirements put together necessitate specialized hardware+software stack. Our mission is to build, own, and manage this specialized stack while still ensuring that infrastructure is performant, efficient, reliable, scalable, and easy to use across different environments.

To achieve our mission we have built massive state-of-the-art High Performance Computing clusters (private HPC cloud), leased capacity from public HPC cloud providers (AWS, GCP) and have built frameworks and libraries to help with running large scale distributed AI training. We are just getting started though and there is a lot more we aim to do. End-to-end researcher workflows is far from ideal and needs to be vastly simplified, moving experiments across different environments (private, FB prod and public cloud) needs to become seamless, efficient storage abstractions/services need to be built, gaps in monitoring, accounting, alerting need to be closed, and the list goes on. In short, today infra and its internal details are visible to our users and we are on a mission to make infra invisible.

Technologies: HPC (GPU, CUDA, NCCL, IB, SLURM), PyTorch, Hydra, ioPath, C++, Python, fairseq, detectron2, vissl