Tackling the Challenges of Imaging-Based AI *for Drug Development*

Adopting AI techniques creates new challenges in data management and curation. Tackling these issues reduces financial and time costs, leading to more efficient drug development.

The last few years have seen a sharp rise in the use of artificial intelligence (AI) and machine learning (ML) within the pharmaceutical sector, including their applications for drug discovery and development.[1] Heavy investment is driving huge growth, with third-party investment in AI-enabled drug discovery more than doubling annually over the last five years.[2]

To fuel this AI revolution, enterprises need vast amounts of data. One way companies can create these large datasets is by leveraging pre-existing data from past clinical studies and in-house research. Using AI and ML techniques, new insights not mined by the original studies can be generated to accelerate drug development.

Since AI algorithms can be highly adaptable to new data and research questions, they can be implemented at many stages of the drug development process, including:

- Drug target discovery
- Drug design to predict structure and interactions
- Drug synthesis, including chemical synthesis and polypharmacology
- Drug repurposing to predict new uses
- Drug screening to predict treatment response and toxicity[3]

However, companies adopting AI into their research face significant data handling and security challenges. Biomedical data, especially imaging data, is prone to inconsistencies in data management and curation. In the past, standardizing complex imaging data to make it AI-ready has taken weeks or months, creating extra financial and time burdens.

At the scale required for AI, even properly formatted data comes with challenges for storage and collaboration. Traditionally, providing access for many users to the same datasets is a complicated process that is difficult to maintain. To successfully collaborate on an AI project, users need to have access to the most up-to-date versions of project data in real time, potentially from different locations and institutions.

AI Is Only Ever as Good *as the Training Data*

The key hurdle for any AI or ML project is curating training datasets. At their core, all AI algorithms can do is find patterns within the data they receive, so the insights they produce are directly impacted by the quality and variety of the data they are trained on. This makes training dataset curation one of the most critical parts of any pipeline.

Non-Standardized Data Is *Still Valuable*

Non-standardized data is one of the most challenging aspects of working with imaging datasets. By the time data reaches a pharma research team, nothing can be done about the way it was collected and recorded, typically leading to many inconsistencies. This situation is often worse when repurposing data that was not collected with quantitative analysis (such as AI applications) in mind. However, the untapped value these datasets still hold cannot be left hidden behind technical challenges, so solutions must be developed.

Centralizing Storage Can *Streamline Processes*

The large quantities of training data needed for AI projects are often distributed across different trials, locations, and servers. Input for the project may be needed across multiple collaborators, and there are likely to be several steps in the analysis pipelines/workflows, including pre- and post-processing. Centralizing all of these datasets can streamline the AI workflow and create consistent curation and analysis pipelines. This initial setup of the data storage and design of a project can have a significant impact on the future financial and time costs it incurs.

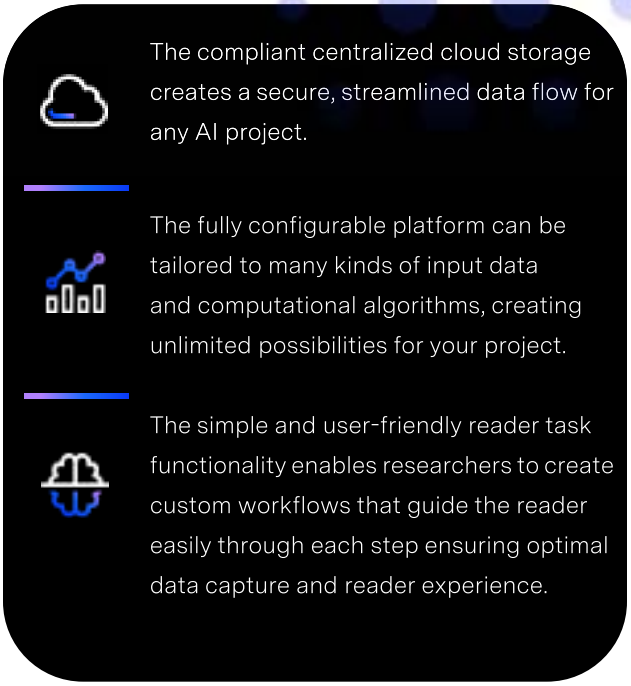
Data Storage and Sharing *Must be Compliant*

A variety of security challenges can arise as companies attempt to introduce and develop new AI workflows. When storing and sharing sensitive medical imaging data, companies must ensure the security and compliance of these processes at all times.[4]

De-identification of medical data is a critical security need for pharmaceutical companies and, for medical imaging data, this comes with extra challenges since identifying details can be embedded within the image itself. For a smooth AI workflow, companies must develop and streamline security solutions for every part of their process–, including the de-ID of all data.

A Single Platform to *Capture, Curate, Compute, and Collaborate*

Flywheel offers centralized storage, automated curation, computational workflows, and secure collaboration within one platform. The platform is scalable, flexible, and fully configurable to accommodate all kinds of imaging data and computational algorithms (including AI approaches), allowing for easy adoption into pharma research workflows.



Curating and *De-identifying Data*

When curating medical imaging data, large file sizes are a key bottleneck. Online file transfers can have long processing times and high failure rates, so researchers often resort to physically moving hard drives. Dr. Geoff Manley, MD, Ph.D., Core Leader of the Transforming Research and Clinical Knowledge in Traumatic Brain Injury (TRACK-TBI) study, used Flyweel's platform to overcome this issue.

All 19 collaborators in the TRACK-TBI project could upload their imaging data to a single cloud location. Perhaps even more importantly, though, the researchers were able to use Flywheel's customizable de-identification templates.

Since each site had its own de-ID needs, the configurable nature of Flywheel's templates allowed the data to be anonymized in a flexible and standardized way across collaborators.

Everything in *One Place*

Flywheel allows you to capture, curate, compute, and collaborate within a single cloud-based platform, providing a streamlined core for AI workflows. Dr. John Garrett, Director of Informatics at the University of Wisconsin Department of Radiology, used Flywheel's technology to collect and curate the chest X-ray data needed to develop an AI COVID-19 diagnostic tool. With Flywheel, Dr. Garrett and his team were able to aggregate data streams from over 35 hospitals and clinics and multiple imaging machines into a single platform. From there, the researchers were able to easily access the data through a web browser interface to audit their training and test data, as well as add their own metadata for custom curation.

> **"If you have the ability to quickly and easily curate and normalize data from many different systems then you're going to be able to do work that's going to be much more impactful." - Dr. John Garrett**

For multicenter research, projects can be designed so that all data and analytic results are available to project members via a secure web browser interface with fully customizable access controls. Whether you are working within your own group or collaboratively, Flywheel is able to make research quicker and simpler than ever before by providing a streamlined, centralized workflow for project data.

Secure Data in *Your Own Way*

A SOC2-certified organization, Flywheel is committed to ensuring data security. Flywheel's experts experts understand the need for individualized and customizable security and compliance, so with Flywheel, you control your own data and user access. The platform can be designed with multiple levels of access for different users to create a secure research environment for pharmaceutical companies and their collaborators.

- Customizable integrated de-identification tools for ingestion of medical imaging data
- GDPR - flexible features that allow for de-ID and personal data removal on request
- 21 CFR Part 11 - functionality for secure data collection, access controls, audit trails, and digital signatures

Automate *Your Workflows*

Once the pipeline has been created on the platform, financial and time costs can be further reduced with automation of curation and analysis pipelines. New data can be fed into already established workflows with ease, giving researchers more time to focus on extracting valuable insights.

Pipeline Development with *Platform Experts*

Flywheel's ethos is to develop the platform in collaboration with clients, ensuring swift and effective data ingestion. Clients can expect a use-case-tailored experience designed to meet their specific project needs and adapted by platform experts wherever necessary.

Flywheel's experts work with each client and project on a use-case basis to develop solutions that meet your unique needs.

AI and machine learning projects can have complicated data curation needs that must be seamlessly incorporated for a successful pipeline.

In the case of the AI COVID-19 diagnostic tool developed by Dr. Garrett's team, the researchers faced data collection from multiple sources and the development of several training datasets. Flywheel was able to help the team to:

- **Ingest** over 10,000 chest X-ray images from patients and healthy controls from 5 hospitals and over 30 clinics
- **Curate** data to exclude patients who did not fit the criteria and classify it into training and test datasets
- **Customize** this classification to ensure data from multiple imaging devices was included within each subset of train and test data to control for variability
- **Export** manifests for training data as .csv or .json files by adding custom metadata tags to images including information about PCR-testing

During setup, Flywheel focuses on data ingestion and standardization. Since non-standardized data creates many challenges for downstream applications, Flywheel's support provides expert advice and implements custom tools to curate data with formatting issues and make it ready for AI and other applications. Clients can discuss their specific needs with platform experts and work together to design custom solutions.

Flywheel's experts remain on hand to develop the rest of the pipeline as needed, helping to overcome individual AI obstacles with the design and implementation of custom algorithms. The goal of the platform and its experts is to create a pipeline that meets the needs of each individual client, empowering them to continue to push the boundaries of what can be done with AI in the pharmaceutical industry.

Collaboration and *Exchange*

To further develop AI pipelines, Flywheel provides a Gear Exchange, where organizations can select and apply containerized algorithms (namely Gears) to any workflow. Gears from Flywheel experts or their clients can be found on the Exchange for use in your research, or your own algorithms can be packaged into custom Gears for effortless use within the platform and future sharing.

Dr. Azeezat Azeez, a postdoctoral scholar with Stanford Psychiatry and Behavioral Sciences, has developed two custom Gears with Flywheel and praised the ease of the process, saying it's "really just packaging what you've already written on your local computer into scripts with a few additional files."

Gears can quickly be swapped and linked together, facilitating the easy revision of drug discovery workflows as technology advances. As Flywheel and its clients continue to develop novel AI approaches, the Gear Exchange will continue to grow, giving clients access to cutting-edge technology trends within imaging-based AI.

In addition to the Gear Exchange, Flywheel has developed a Data Exchange for its clients. Publicly available imaging datasets are integral to AI research for drug development; however, the ingestion of this data is cumbersome and vulnerable to formatting issues. The Data Exchange brings these pre-curated datasets into the platform and allows clients to access them without ingestion and curation challenges.

Automated data ingestion and standardization make the setup process quicker and easier, reducing the initial barriers to project startup.

Continued support for pipeline development allows clients to more easily develop their ideal workflows and reduce problem points in their projects.

Gears and the Data Exchange create faster and easier access to ready-to-use data and algorithms, all within a single platform.

Reduced Costs and *Quicker Development*

By adopting Flywheel's platform and utilizing their professional, supportive services, companies can tackle the major challenges of AI approaches in drug development, including:

- Data ingestion
- Data standardizing and formatting
- Data labeling and annotation
- Creation of automated analysis pipelines
- Data storage and collaboration
- Security, access, and compliance

Flywheel's services allow you to overcome the key barriers to AI project momentum while freeing up valuable time and resources. The platform can help lower costs and increase research output to speed up drug development. Start your journey to streamlined analysis pipelines and accelerated drug development today.

Learn more about how Flywheel can work for you by visiting **www.flywheel.io** *or* **booking a demo.**

Flywheel is the pioneering medical imaging data and AI platform powering healthcare innovation through streamlined medical imaging data management, curation and analysis. Flywheel helps organizations turn complex imaging data into analysis-ready datasets for accelerated research and AI development. Flywheel offers comprehensive solutions for pharma companies, providers, payers, system integrators, AI developers and academic medical centers to get optimum value out of their data assets. Flywheel is an Invenshure-founded company headquartered in Minneapolis, with offices in the Bay Area, St. Louis, and Budapest. For more information on our mission and products, visit:

www.flywheel.io
Follow us on Twitter and LinkedIn

References:

1. Sahu A, Mishra J, Kushwaha N. Artificial Intelligence (AI) in Drugs and Pharmaceuticals. Comb Chem High Throughput Screen. 2022;25(11):1818-1837.

2. Adopting AI in Drug Discovery. BCG Global. Published March 23, 2022. Accessed June 9, 2023. https://www.bcg.com/publications/2022/adopting-ai-in-pharmaceutical-discovery

3. Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. Drug Discov Today. 2021;26(1):80-93. doi:10.1016/j.drudis.2020.10.010

4. Mireku A. 2022 in review: Regulation starts to catch up with AI in pharma. Pharmaceutical Technology. Published December 22, 2022. Accessed May 17, 2023. https://www.pharmaceutical-technology.com/features/2022-in-review-regulation-starts-to-catch-up-with-ai-in-pharma/