UNITED STATES
SECURITIES AND EXCHANGE COMMISSION
Washington, D.C. 20549

FORM 6-K

REPORT OF FOREIGN PRIVATE ISSUER PURSUANT TO RULE 13a-16 OR 15d-16
UNDER THE SECURITIES EXCHANGE ACT OF 1934

FOR THE MONTH OF OCTOBER 2024

COMMISSION FILE NUMBER 001-39081

BioNTech SE
(Translation of registrant's name into English)

An der Goldgrube 12
D-55131 Mainz
Germany
+49 6131-9084-0
(Address of principal executive offices)

Indicate by check mark whether the registrant files or will file annual reports under cover Form 20-F or Form 40-F: Form 20-F ☒ Form 40-F ☐

Indicate by check mark if the registrant is submitting the Form 6-K in paper as permitted by Regulation S-T Rule 101(b)(1): ☐

Indicate by check mark if the registrant is submitting the Form 6-K in paper as permitted by Regulation S-T Rule 101(b)(7): ☐

DOCUMENTS INCLUDED AS PART OF THIS FORM 6-K

On October 1, 2024, BioNTech SE (the "Company"), alongside its artificial intelligence ("AI") subsidiary InstaDeep Ltd., is presenting an overview of its AI approach during an edition of the Company's Innovation Series, AI Day. The press release and presentation are attached as Exhibits 99.1 and 99.2, respectively.

SIGNATURE

Pursuant to the requirements of the Exchange Act, the registrant has duly caused this report to be signed on its behalf by the undersigned, thereunto duly authorized.

BioNTech SE

By: /s/ Jens Holstein
Name: Jens Holstein
Title: Chief Financial Officer

By: /s/ Dr. Sierk Poetting
Name: Dr. Sierk Poetting
Title: Chief Operating Officer

Date: October 1, 2024

EXHIBIT INDEX

Exhibit	Description of Exhibit
99.1	BioNTech Highlights AI Capabilities and R&D Use Cases at Inaugural AI Day
99.2	Presentation

BioNTech Highlights AI Capabilities and R&D Use Cases at Inaugural AI Day

- *Provides updates on BioNTech's strategy to scale and deploy AI-capabilities across the immunotherapy pipeline*
- *Highlights InstaDeep's new near exascale supercomputer, Kyber, with the aim of enabling BioNTech to compute at scale*
- *Unveils novel AI Bayesian Flow Network ("BFN") models for protein sequence generation*
- *Highlights progress in deploying AI across BioNTech's immunotherapy pipeline, including in immunohistochemistry, DNA/RNA sequencing, proteomics, protein design, and lab functional validation*

LONDON, United Kingdom, October 1, 2024 (GLOBE NEWSWIRE) -- BioNTech SE (Nasdaq: BNTX, "BioNTech" or "the Company"), alongside its artificial intelligence ("AI") subsidiary InstaDeep Ltd. ("InstaDeep"), today presents an overview of its AI approach during an edition of the Company's Innovation Series, AI Day.

"At BioNTech, we are at the forefront of integrating advanced AI to revolutionize individualized medicine. With our in-house AI specialist, InstaDeep, we are pioneering the use of artificial intelligence to develop personalized vaccines and targeted therapies," said **Prof. Ugur Sahin, M.D., CEO and Co-Founder of BioNTech**. "By introducing state-of-the-art technologies such as our BFN generative protein model and incorporating AI capabilities across our immunotherapy pipeline, we are unlocking the full potential of AI to deliver innovative vaccines and cancer treatments to patients worldwide".

"BioNTech and InstaDeep, as a biotechnology powerhouse with a dedicated AI unit, are uniquely positioned at the intersection of biotechnology and AI," said **Karim Beguir, CEO and Co-Founder of InstaDeep, a BioNTech company**. "Working closely together and combining expertise from AI research, high performance computing, software and biology is accelerating innovation. A key focus of this collaboration is our DeepChain™ multiomics design platform. DeepChain™ is now open for external partnerships, after successful application to several projects, including the mRNA-encoded antibody RiboMab™ platform. We are excited to harness breakthroughs in both AI and biotechnology synergistically.''

As part of the event, BioNTech will showcase the Company's approach to AI capability scaling and deployment across BioNTech's pipeline. These updates will cover the introduction of a new near exascale supercomputer, the launch of a novel BFN generative model, and multiple updates on the deployment of AI across BioNTech's immunotherapy pipeline.

The live webcast of the event will be available via this link and will begin at 3:00 pm CEST (2:00 pm BST). A replay of the webcast will be available shortly after the event's conclusion and archived on BioNTech's website for one year.

About BioNTech
Biopharmaceutical New Technologies (BioNTech) is a global next generation immunotherapy company pioneering novel therapies for cancer and other serious diseases. BioNTech exploits a wide array of computational discovery and therapeutic drug platforms for the rapid development of novel biopharmaceuticals. Its broad portfolio of oncology product candidates includes individualized and off-the-shelf mRNA-based therapies, innovative chimeric antigen receptor (CAR) T cells, several protein-based therapeutics, including bispecific immune checkpoint modulators, targeted cancer antibodies and antibody-drug conjugate (ADC) therapeutics, as well as small molecules. Based on its deep expertise in mRNA vaccine development and in-house manufacturing capabilities, BioNTech and its collaborators are developing multiple mRNA vaccine candidates for a range of infectious diseases alongside its diverse oncology pipeline. BioNTech has established a broad set of relationships with

multiple global and specialized pharmaceutical collaborators, including Biotheus, DualityBio, Fosun Pharma, Genentech, a member of the Roche Group, Genevant, Genmab, MediLink, OncoC4, Pfizer and Regeneron.

For more information, please visit www.biontech.com.

Forward-Looking Statements
This press release contains forward-looking statements within the meaning of the Private Securities Litigation Reform Act of 1995, as amended, including, but not limited to, statements concerning the impact of, and expectations regarding, BioNTech's AI capabilities and future research and development activities. In some cases, forward-looking statements can be identified by terminology such as "will," "may," "should," "expects," "intends," "plans," "aims," "anticipates," "believes," "estimates," "predicts," "potential," "continue," or the negative of these terms or other comparable terminology, although not all forward-looking statements contain these words. The forward-looking statements in this press release are based on BioNTech's current expectations and beliefs of future events, and are neither promises nor guarantees. You should not place undue reliance on these forward-looking statements because they involve known and unknown risks, uncertainties, and other factors, many of which are beyond BioNTech's control and which could cause actual results to differ materially and adversely from those expressed or implied by these forward-looking statements. These risks and uncertainties include, but are not limited to: the uncertainties inherent in research and development,; competition from other product candidates BioNTech's and its counterparties' ability to manage and source necessary energy resources; BioNTech's ability to identify research opportunities and discover and develop investigational medicines; the ability and willingness of BioNTech's third-party collaborators to continue research and development activities relating to BioNTech's development candidates and investigational medicines; BioNTech's ability to manage its development; regulatory developments in the United States and other countries; BioNTech's ability to effectively scale its production capabilities and manufacture its products and product candidates; risks relating to the global financial system and markets; and other factors not known to BioNTech at this time.

You should review the risks and uncertainties described under the heading "Risk Factors" in BioNTech's Report on Form 6-K for the period ended June 30, 2024 and in subsequent filings made by BioNTech with the SEC, which are available on the SEC's website at www.sec.gov. These forward-looking statements speak only as of the date hereof. Except as required by law, BioNTech disclaims any intention or responsibility for updating or revising any forward-looking statements contained in this press release in the event of new information, future developments or otherwise.

CONTACTS

Investor Relations
Michael Horowicz
Investors@biontech.de

Media Relations
Jasmina Alatovic
+49 (0)6131 9084 1513
Media@biontech.de

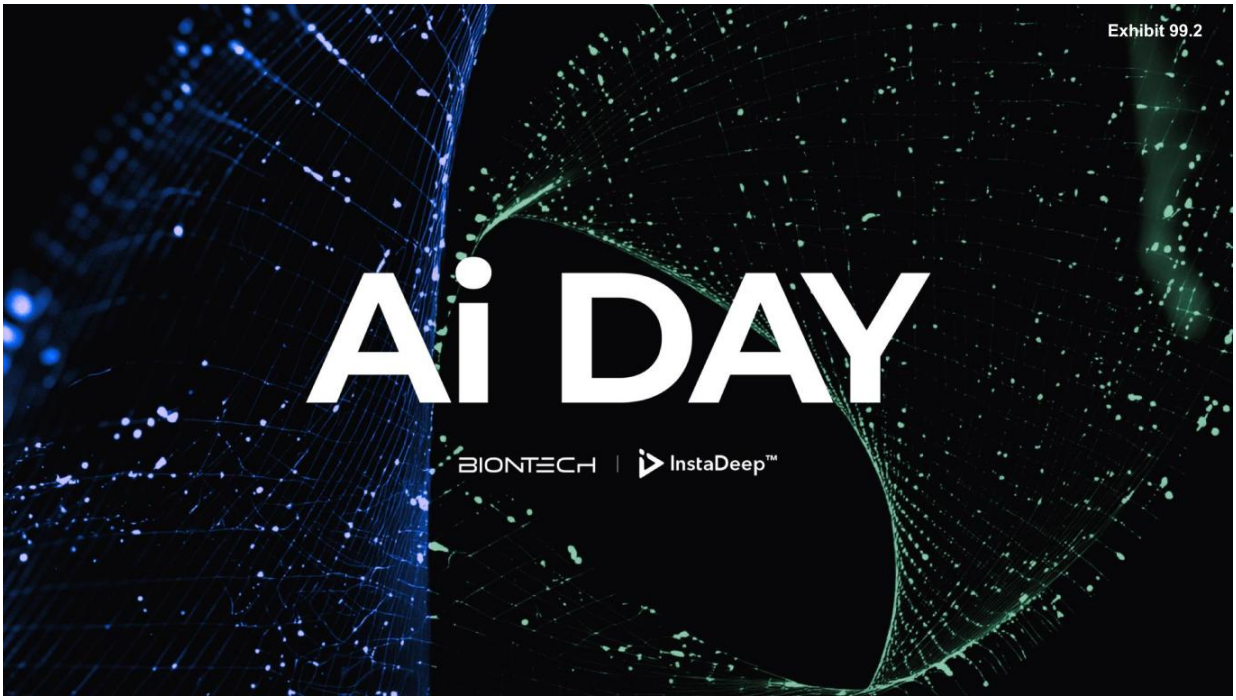


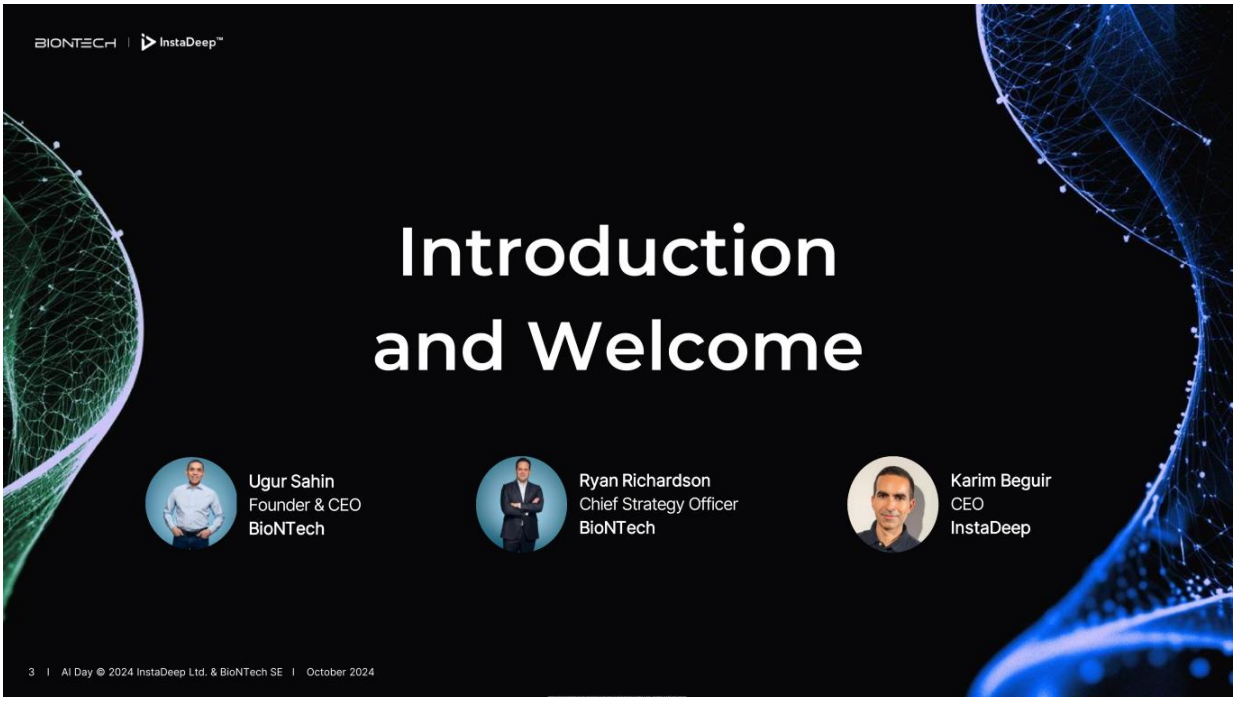
Exhibit 99.2
Ai DAY
BIONTECH
InstaDeep™

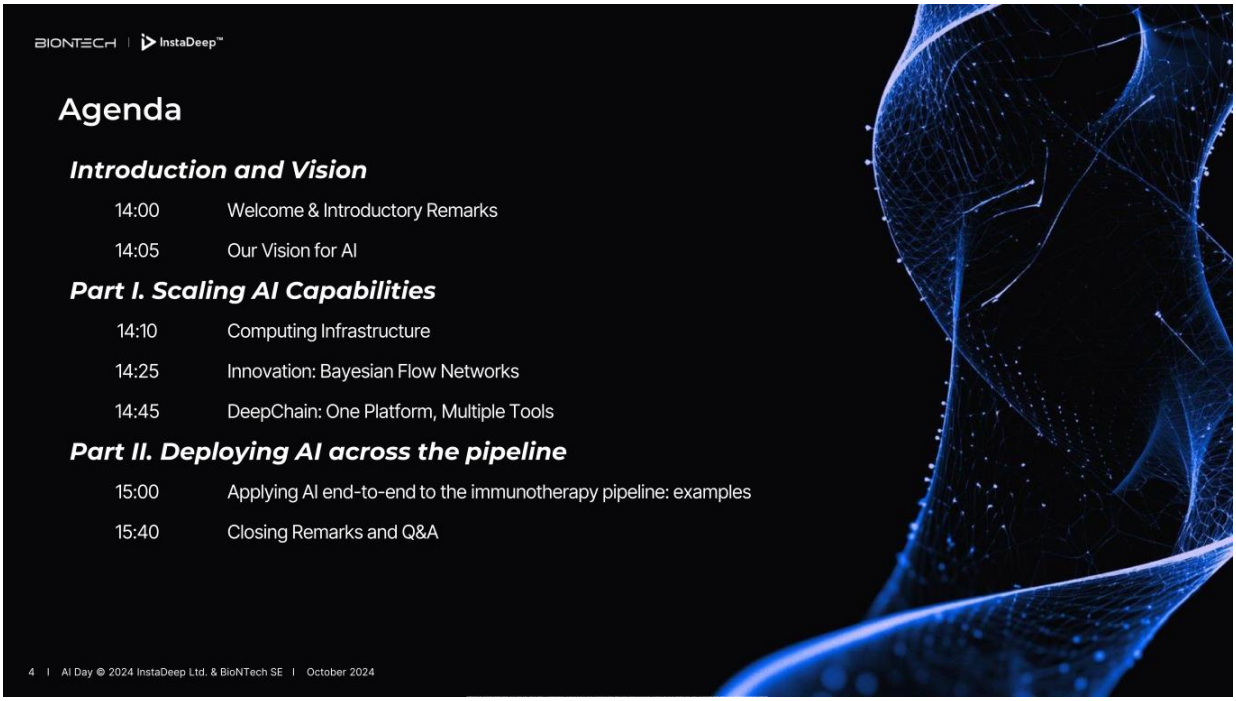
BIONTECH | InstaDeep™

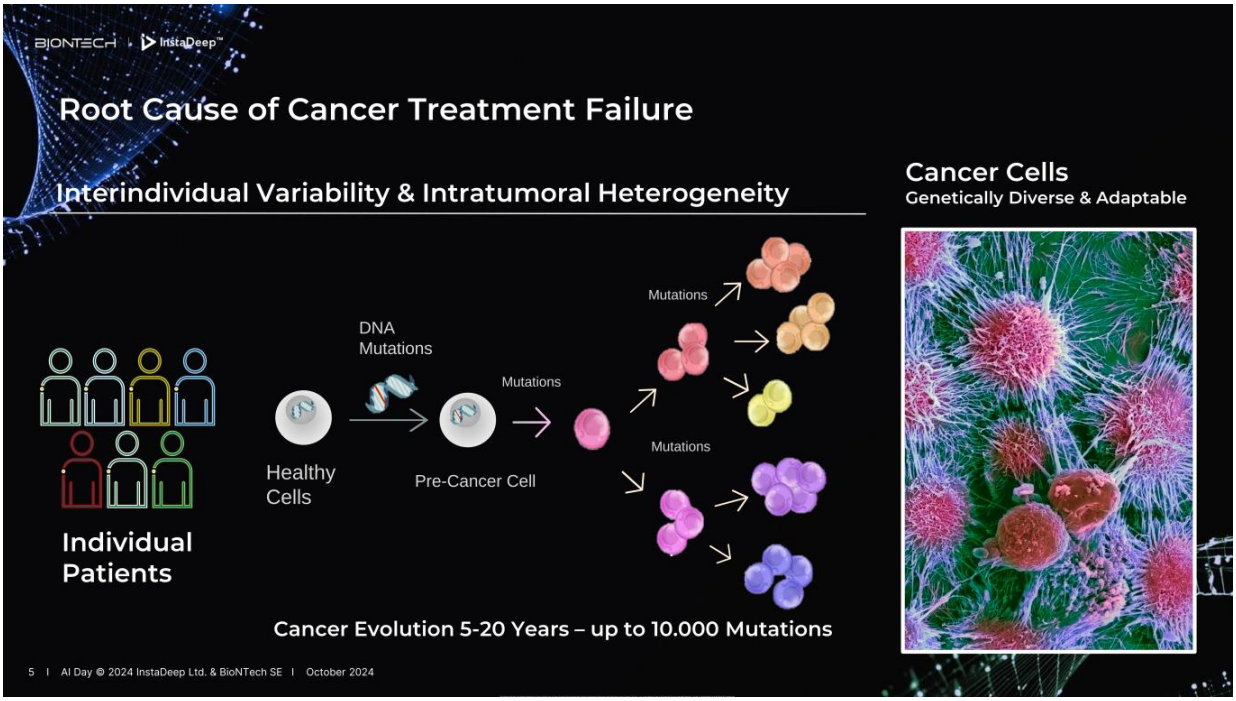
This Slide Presentation Includes Forward-Looking Statements

This presentation contains forward-looking statements within the meaning of the Private Securities Litigation Reform Act of 1995, as amended. In some cases, forward-looking statements can be identified by terminology such as "will," "may," "should," "expects," "intends," "plans," "aims," "anticipates," "believes," "estimates," "predicts," "potential," "continue," or the negative of these terms or other comparable terminology, although not all forward-looking statements contain these words. The forward-looking statements in this presentation are neither promises nor guarantees, and you should not place undue reliance on these forward-looking statements because they involve known and unknown risks, uncertainties, and other factors, many of which are beyond BioNTech's control and which could cause actual results to differ materially from those expressed or implied by these forward-looking statements. You should review the risks and uncertainties described under the heading "Risk Factors" in BioNTech's Quarterly Report on Form 6-K for the period ended June 30, 2024 and in subsequent filings made by BioNTech with the SEC, which are available on the SEC's website at https://www.sec.gov/. Except as required by law, BioNTech disclaims any intention or responsibility for updating or revising any forward-looking statements contained in this presentation in the event of new information, future developments or otherwise. These forward-looking statements are based on BioNTech's current expectations and speak only as of the date hereof.

Furthermore, certain statements contained in this presentation relate to or are based on studies, publications, surveys and other data obtained from third-party sources and BioNTech's own internal estimates and research. While BioNTech believes these third-party sources to be reliable as of the date of this presentation, it has not independently verified, and makes no representation as to the adequacy, fairness, accuracy or completeness of, any information obtained from third-party sources. In addition, any market data included in this presentation involves assumptions and limitations, and there can be no guarantee as to the accuracy or reliability of such assumptions. While BioNTech believes its own internal research is reliable, such research has not been verified by any independent source. In addition, BioNTech is the owner of various trademarks, trade names and service marks that may appear in this presentation. Certain other trademarks, trade names and service marks appearing in this presentation are the property of third parties. Solely for convenience, the trademarks and trade names in this presentation may be referred to without the ® and TM symbols, but such references should not be construed as any indicator that their respective owners will not assert, to the fullest extent under applicable law, their rights thereto.

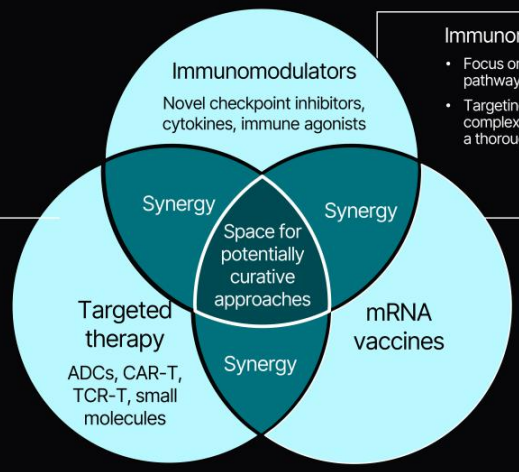

BIONTECH | InstaDeep™
Introduction and Welcome
Ugur Sahin
Founder & CEO
BioNTech
Ryan Richardson
Chief Strategy Officer
BioNTech
Karim Beguir
CEO
InstaDeep
3 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
Agenda
Introduction and Vision
14:00 Welcome & Introductory Remarks
14:05 Our Vision for AI
Part I. Scaling AI Capabilities
14:10 Computing Infrastructure
14:25 Innovation: Bayesian Flow Networks
14:45 DeepChain: One Platform, Multiple Tools
Part II. Deploying AI across the pipeline
15:00 Applying AI end-to-end to the immunotherapy pipeline: examples
15:40 Closing Remarks and Q&A
4 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH
InstaDeep
Root Cause of Cancer Treatment Failure
Interindividual Variability & Intratumoral Heterogeneity
Cancer Cells
Genetically Diverse & Adaptable
DNA
Mutations
Mutations
Mutations
Healthy
Cells
Pre-Cancer Cell
Individual
Patients
Cancer Evolution 5-20 Years – up to 10.000 Mutations
5 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

BIONTECH | InstaDeep™

Towards a Potentially Curative Approach to Cancer Based on Multiple Modalities and Differentiated Novel/Novel Therapeutic Combinations



Immunomodulators

- Focus on the most relevant and crucial IO pathways
- Targeting different complementary players in the complex cancer immunity cycle aiming to achieve a thorough and durable anti-tumor effect

Targeted therapy

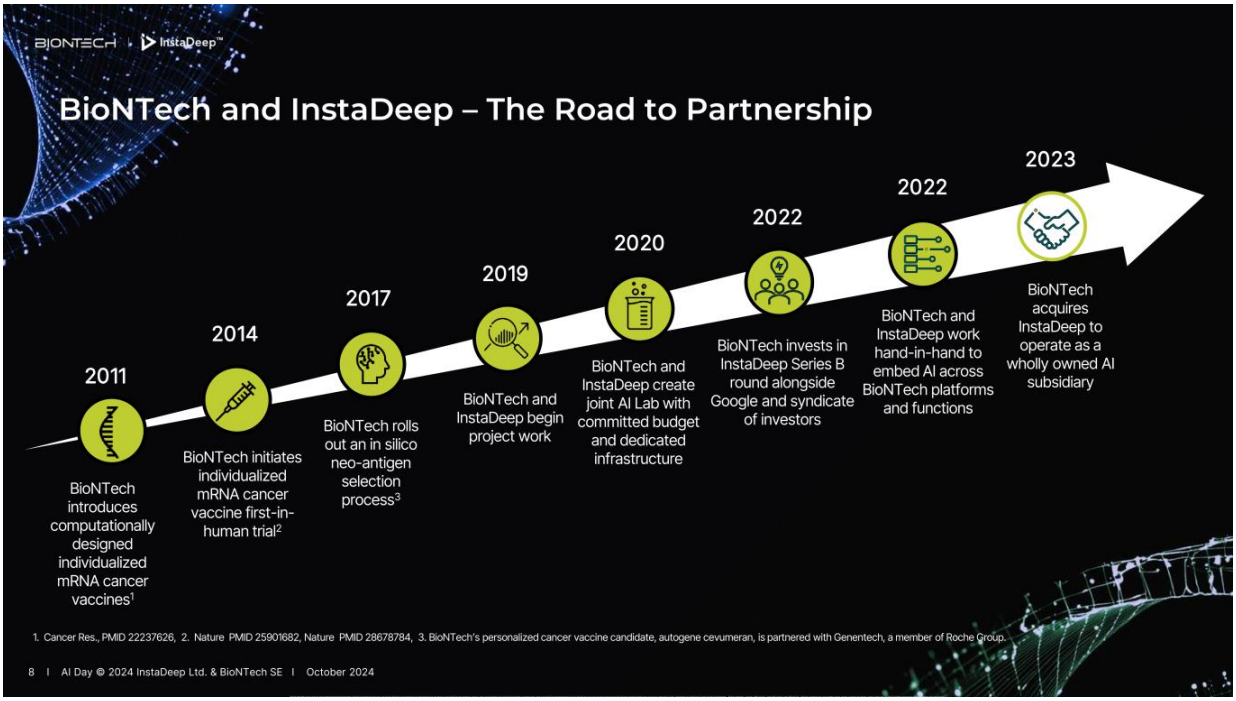
- Could rapidly reduce tumor burden
- Designed with the aim to have clinical efficacy across the entire disease continuum including late lines

mRNA cancer vaccines

- Could eliminate polyclonal residual disease with individualized vaccines for potential long-term impact
- Polyspecific activity by targeting multiple antigens at once

ADC = antibody-drug conjugate; CAR = chimeric antigen receptor; TCR-T = T-cell receptor engineered T cell; IO = immune oncology.


BIONTECH | InstaDeep™
BioNTech:
A pioneer in ML from day one
7 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH
InstaDeep
BioNTech and InstaDeep – The Road to Partnership
2011
BioNTech introduces computationally designed individualized mRNA cancer vaccines[1]
2014
BioNTech initiates individualized mRNA cancer vaccine first-in-human trial[2]
2017
BioNTech rolls out an in silico neo-antigen selection process[3]
2019
BioNTech and InstaDeep begin project work
2020
BioNTech and InstaDeep create joint AI Lab with committed budget and dedicated infrastructure
2022
BioNTech invests in InstaDeep Series B round alongside Google and syndicate of investors
2022
BioNTech and InstaDeep work hand-in-hand to embed AI across BioNTech platforms and functions
2023
BioNTech acquires InstaDeep to operate as a wholly owned AI subsidiary
1. Cancer Res., PMID 22237626, 2. Nature PMID 25901682, Nature PMID 28678784, 3. BioNTech's personalized cancer vaccine candidate, autogene cevumeran, is partnered with Genentech, a member of Roche Group.
8 I AI Day © 2024 InstaDeep Ltd. & BioNTech SE I October 2024

BIONTECH | InstaDeep™

Two Companies: One Mission

BioNTech: >6,800[1] Employees	HQ: Mainz, Germany	Developing medicines to fight cancer, infectious diseases and other serious diseases.
InstaDeep: >370[1] Employees	HQ: London, UK	Focused on productizing disruptive AI innovation

Our Goal: Building a leading AI-first, personalized immunotherapy platform

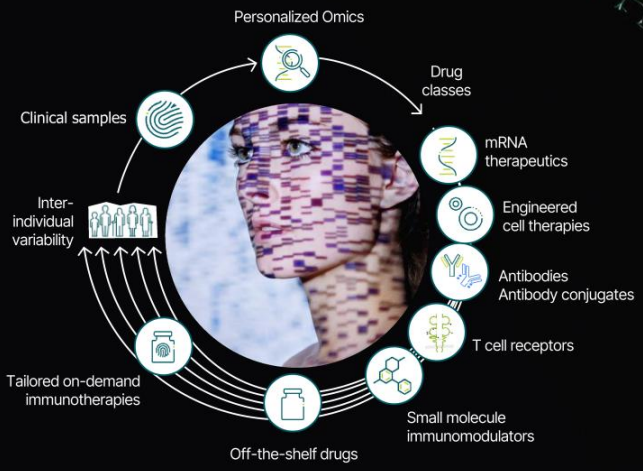
(and leveraging the breakthroughs obtained in the process)

1. As of 30 June 2024

BIONTECH | InstaDeep™

Charting the Course for Tomorrow's Personalized Medicine

- Deep genomics & immunology expertise to analyse patient data
- Individualized treatment platforms to address inter-individual variability
- AI & digitally-integrated target & drug discovery and development
- In-house manufacturing



BIONTECH | InstaDeep™

Our Goal: Deploying AI end-to-end in our immunotherapy pipeline

AI-first Immunotherapy Platform

1. Immunohistochemistry → 2. DNA/RNA sequencing → 3. Proteomics → 4. Protein Design → 5. Lab Functional Validation

Histology

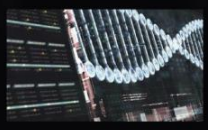
AI computer vision to improve speed and accuracy of tissue labeling

AI Vision

DNA

Sequence analysis and personalized genomic annotations

DNA LLMs



Proteomics

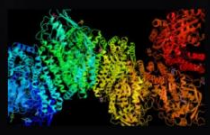
Leveraging AI for target discovery and analysis of the immunological landscape

Protein LLMs

Protein Design

Developing assets for immunotherapy modalities (antibodies, cytokines, TCRs)

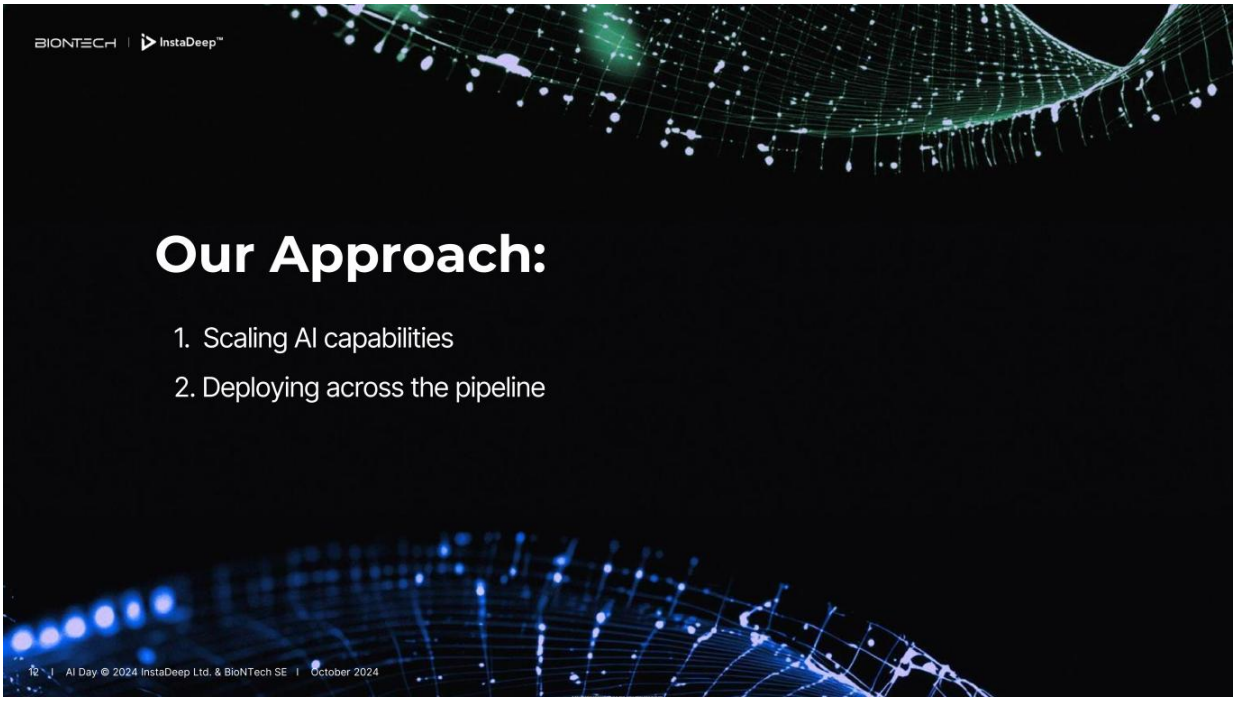
Protein LLMs



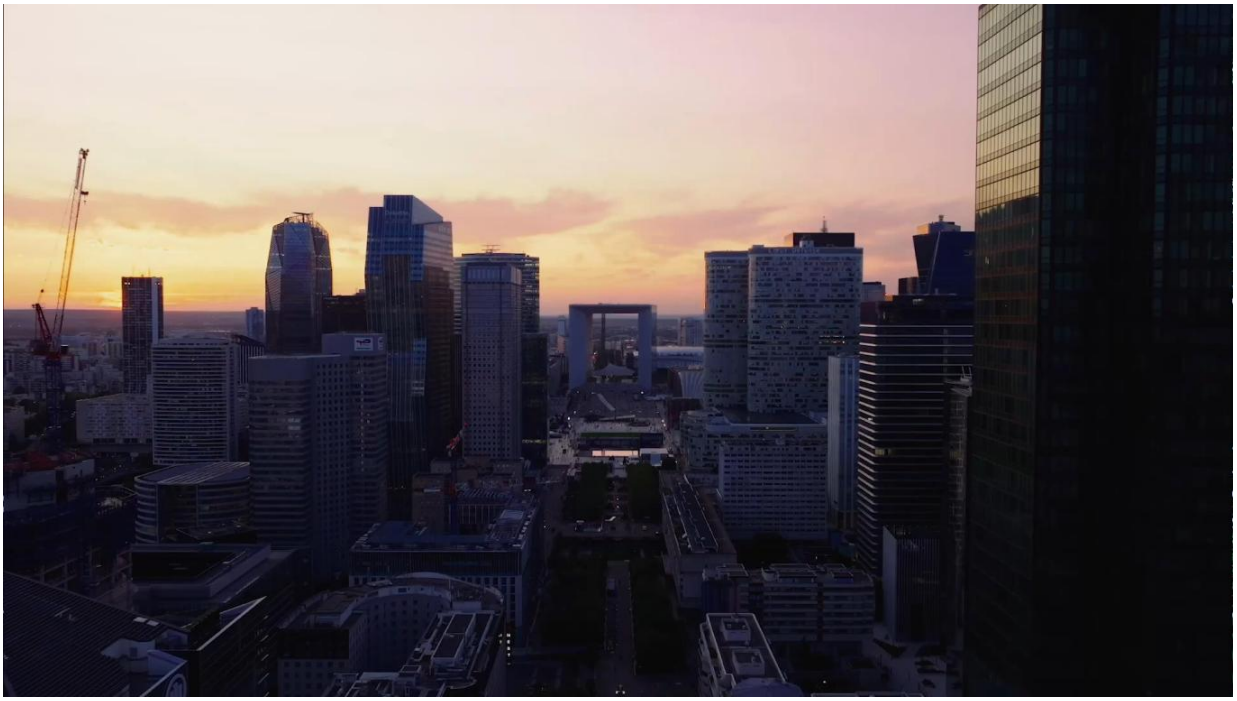
Lab

Instrument automation and quality control

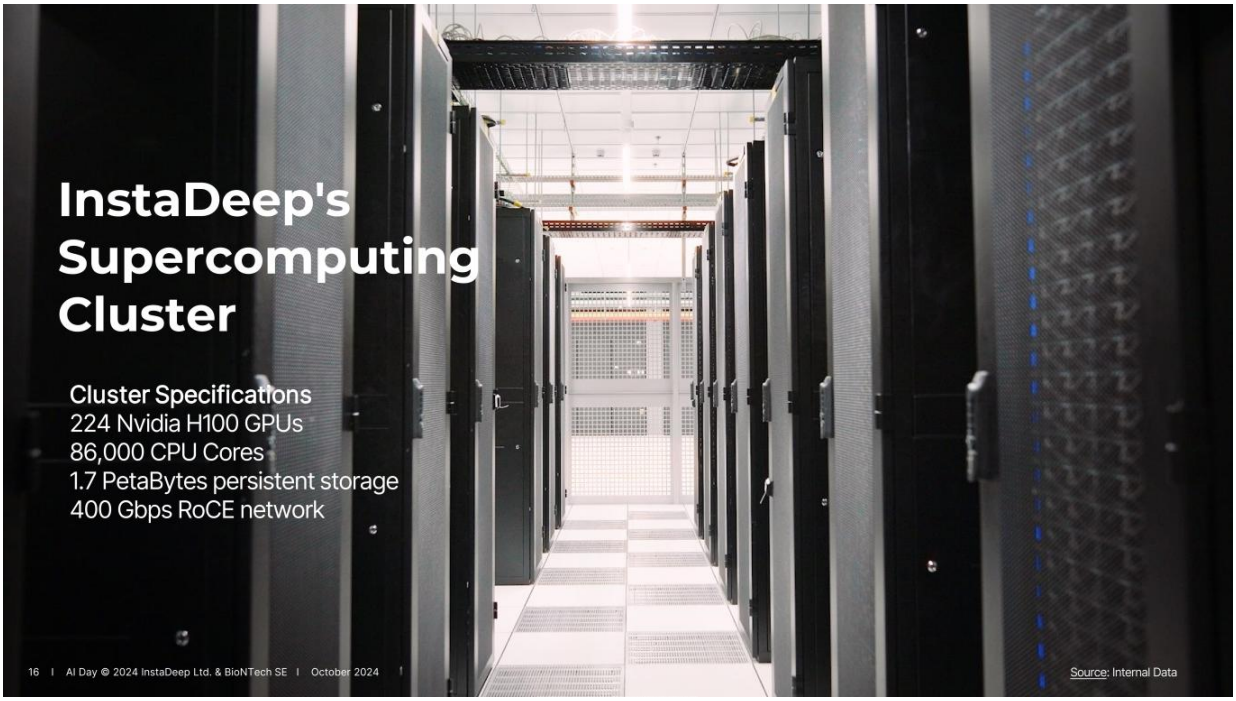
AI Agent

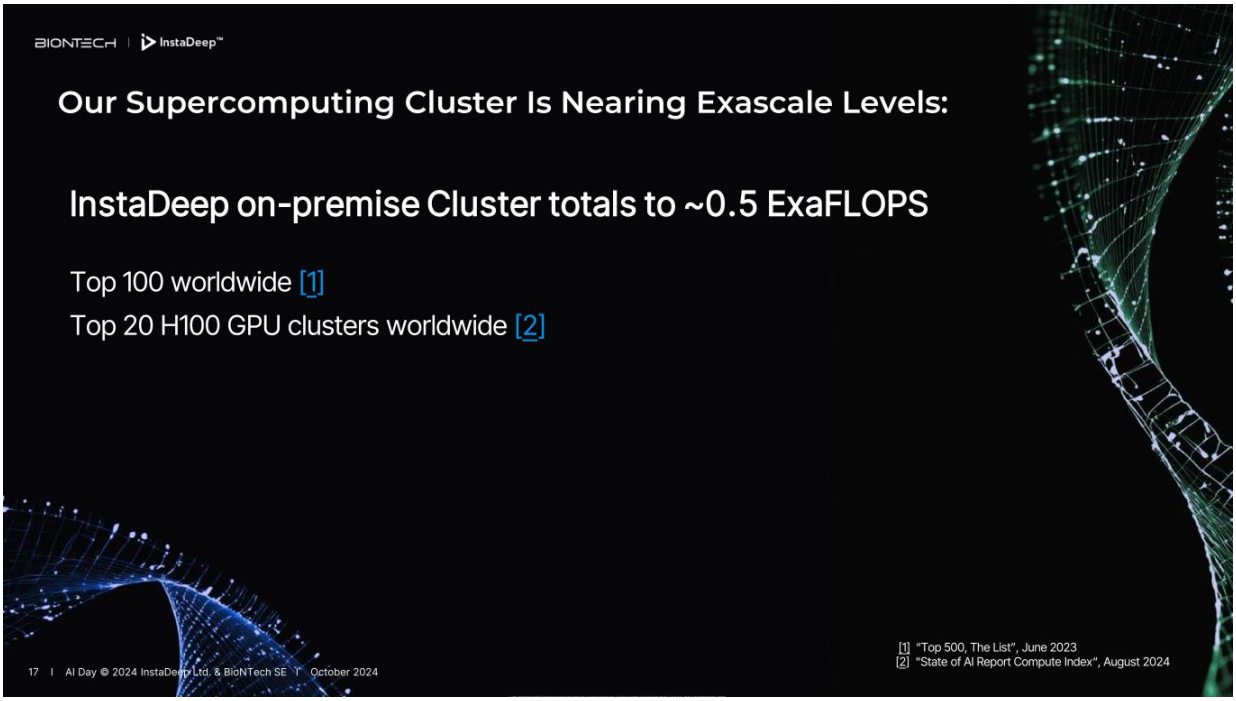

BIONTECH | InstaDeep™
Our Approach:
1. Scaling AI capabilities
2. Deploying across the pipeline
12 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
Part I.
Scaling AI Capabilities
13 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024




BIONTECH | InstaDeep™
Computing
Infrastructure
15 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

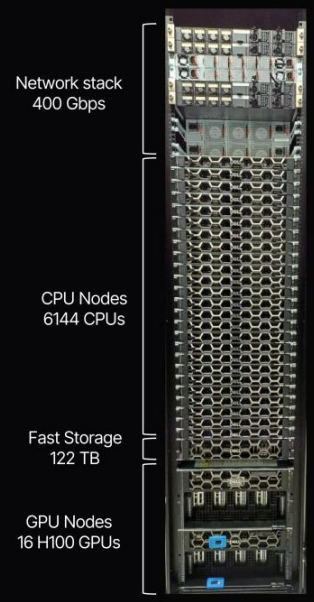

InstaDeep's Supercomputing Cluster
Cluster Specifications
224 Nvidia H100 GPUs
86,000 CPU Cores
1.7 PetaBytes persistent storage
400 Gbps RoCE network
16 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024 |
Source: Internal Data


BIONTECH | InstaDeep™
Our Supercomputing Cluster Is Nearing Exascale Levels:
InstaDeep on-premise Cluster totals to ~0.5 ExaFLOPS
Top 100 worldwide [1]
Top 20 H100 GPU clusters worldwide [2]
17 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024
[1] "Top 500, The List", June 2023
[2] "State of AI Report Compute Index", August 2024

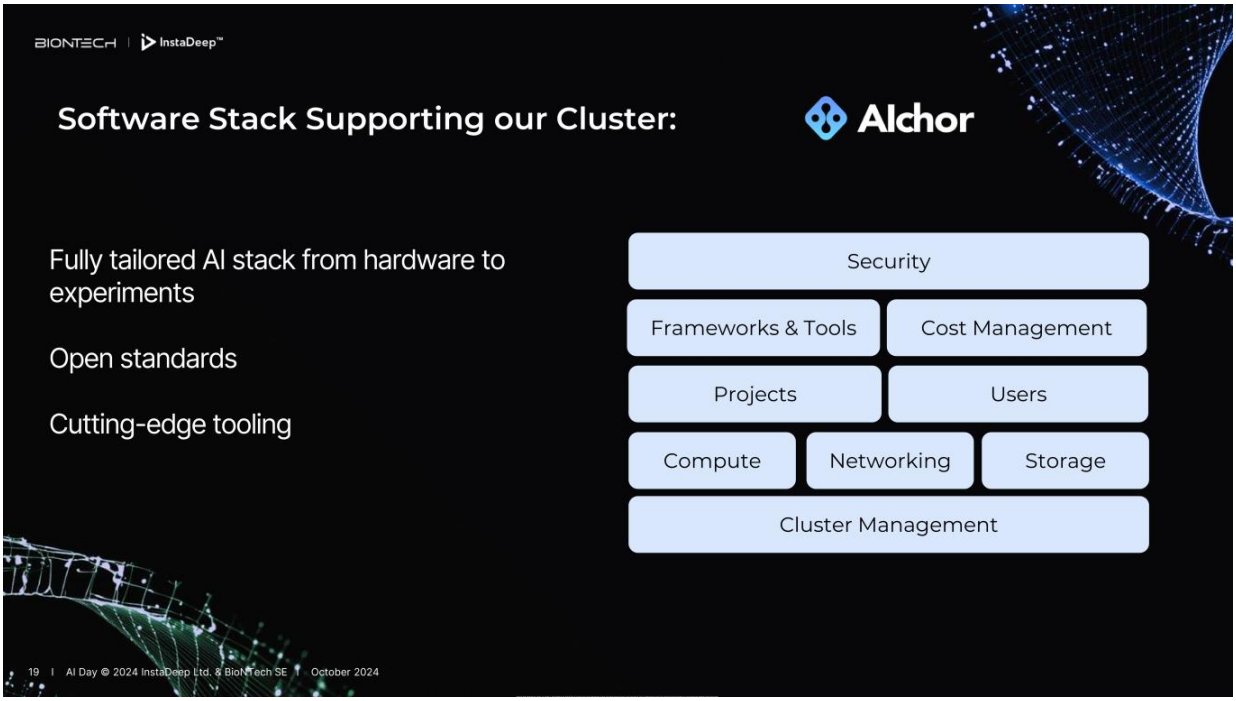
BIONTECH | InstaDeep™

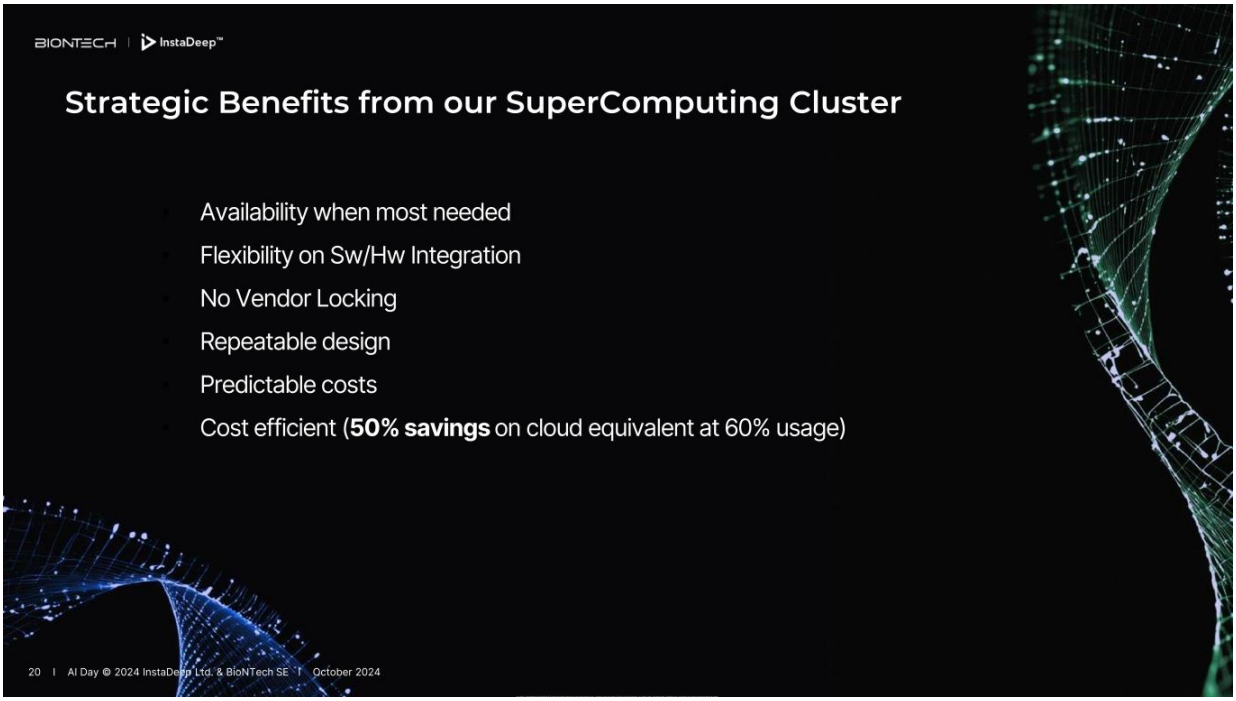
Advanced In-House Rack Design

- Easy to expand with modular nodes
- Consistent performance, cost, power, cooling
- Optimized for large-scale AI workloads
- Simplified management with consistent architecture
- Minimize expenses with standard design

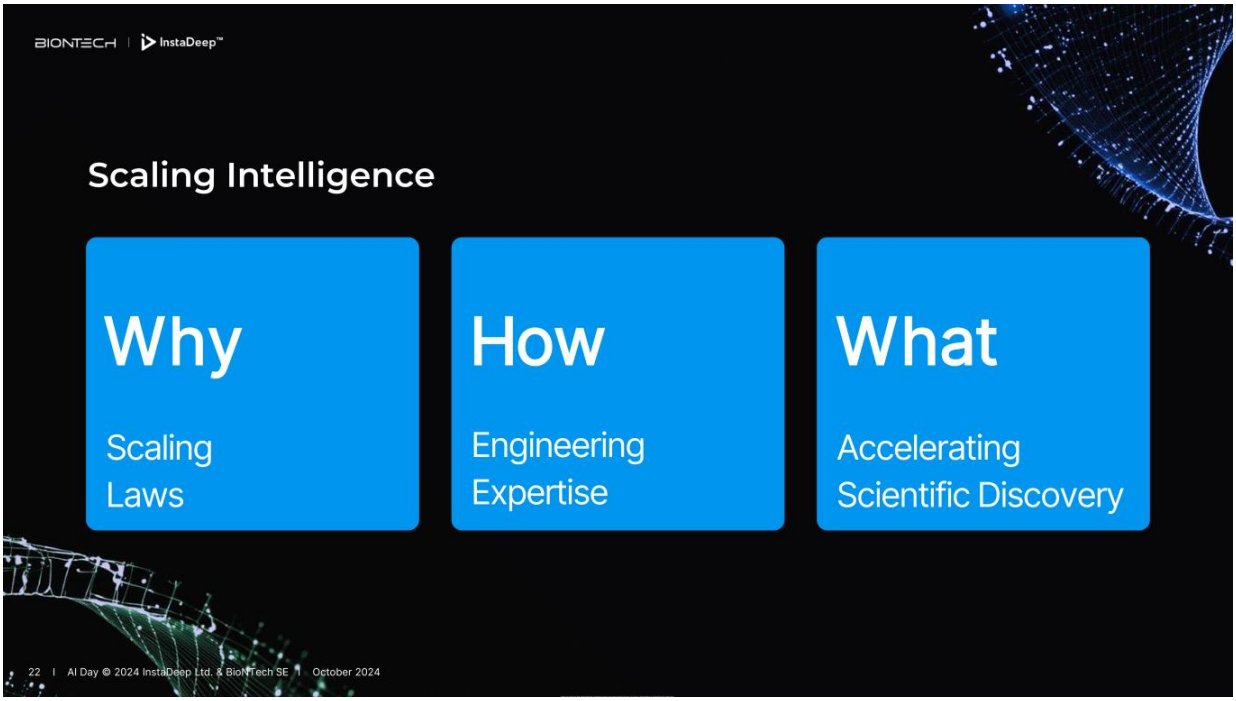


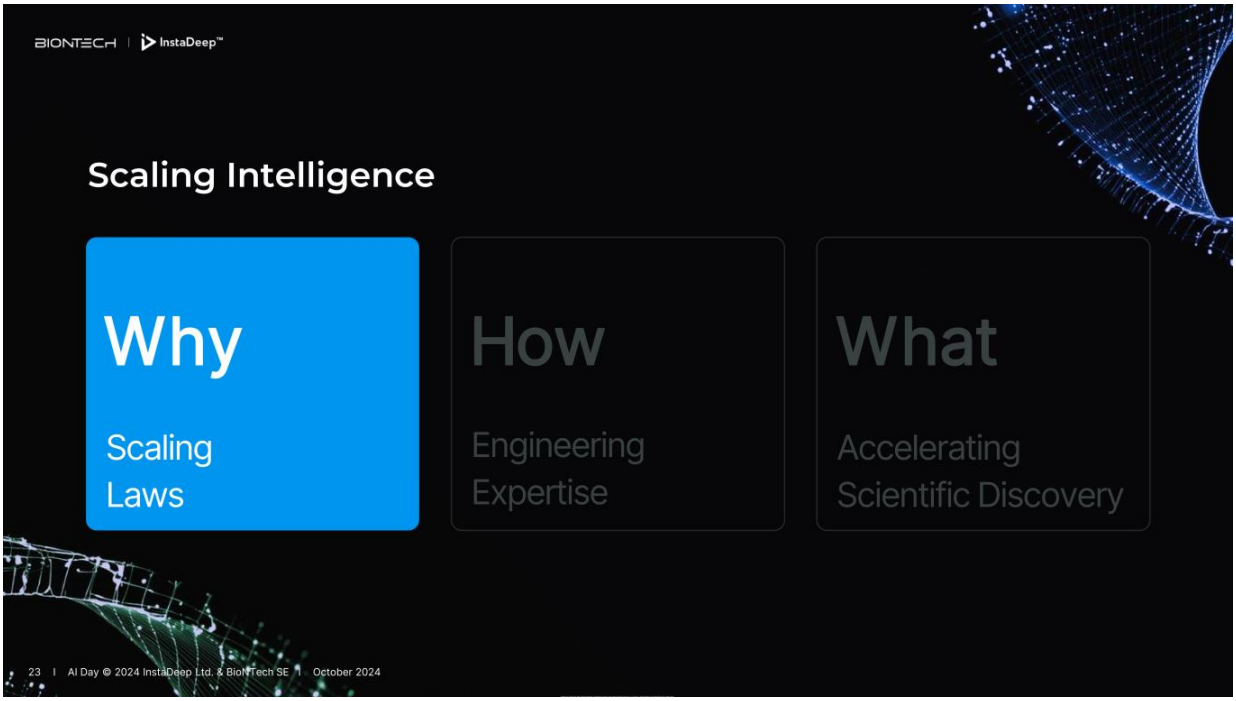
Source: Internal Data


BIONTECH | InstaDeep™
Software Stack Supporting our Cluster:
Alchor
Fully tailored AI stack from hardware to experiments
Open standards
Cutting-edge tooling
Security
Frameworks & Tools
Cost Management
Projects
Users
Compute
Networking
Storage
Cluster Management
19 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
Strategic Benefits from our SuperComputing Cluster
Availability when most needed
Flexibility on Sw/Hw Integration
No Vendor Locking
Repeatable design
Predictable costs
Cost efficient (50% savings on cloud equivalent at 60% usage)
20 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
Scaling Intelligence
21 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

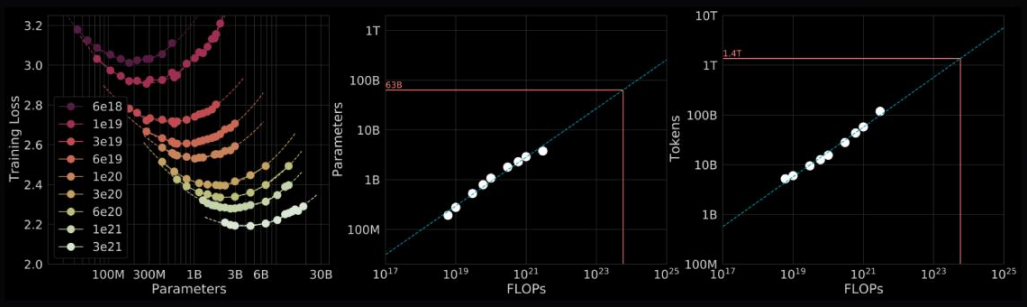

BIONTECH | InstaDeep™
Scaling Intelligence
Why
Scaling Laws
How
Engineering Expertise
What
Accelerating Scientific Discovery
22 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
Scaling Intelligence
Why
Scaling Laws
How
Engineering Expertise
What
Accelerating Scientific Discovery
23 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

BIONTECH | InstaDeep™

Scaling Laws

Performance of Large Language Models (LLMs) is a smooth, well-behaved and predictable function of the **number of parameters** of your model, the **amount of data** used to train it, and **computing resources**.

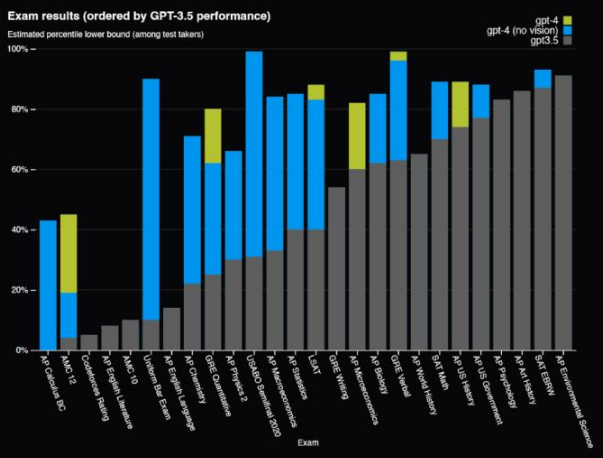


Source: Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D.D.L., Hendricks, L.A., Welbl, J., Clark, A. and Hennigan, T., 2022. Training compute-optimal large language models.

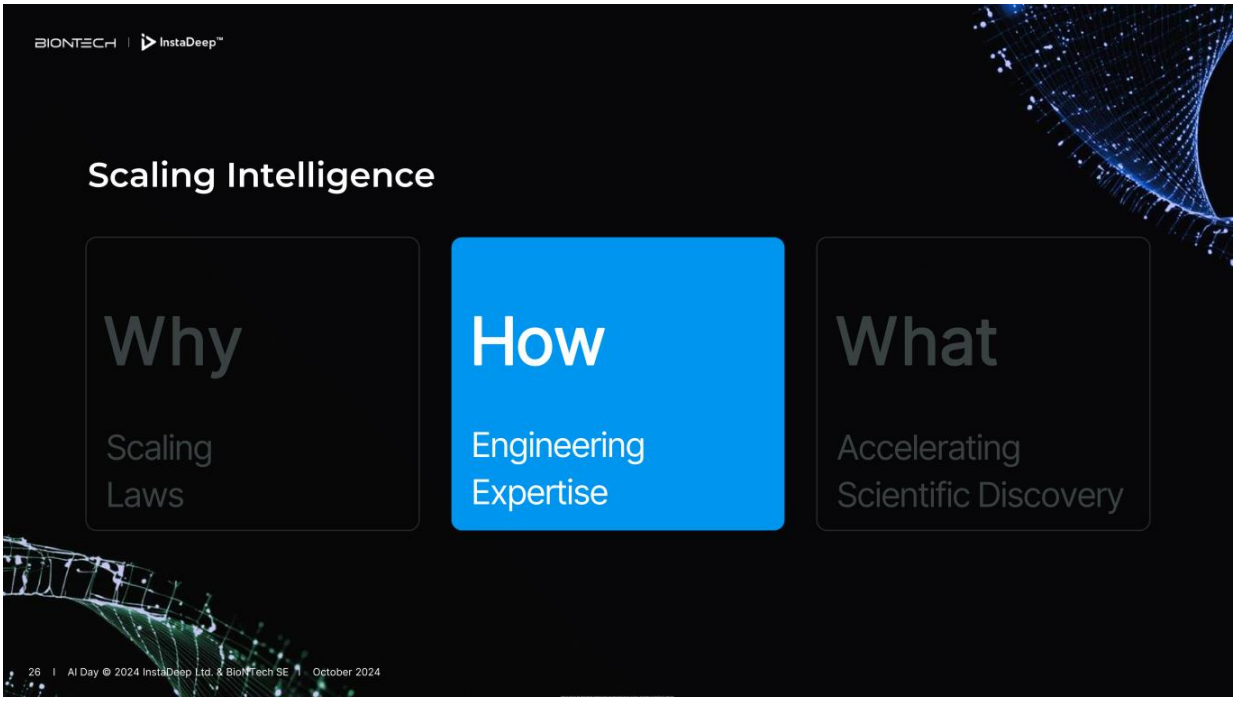
BIONTECH | InstaDeep™

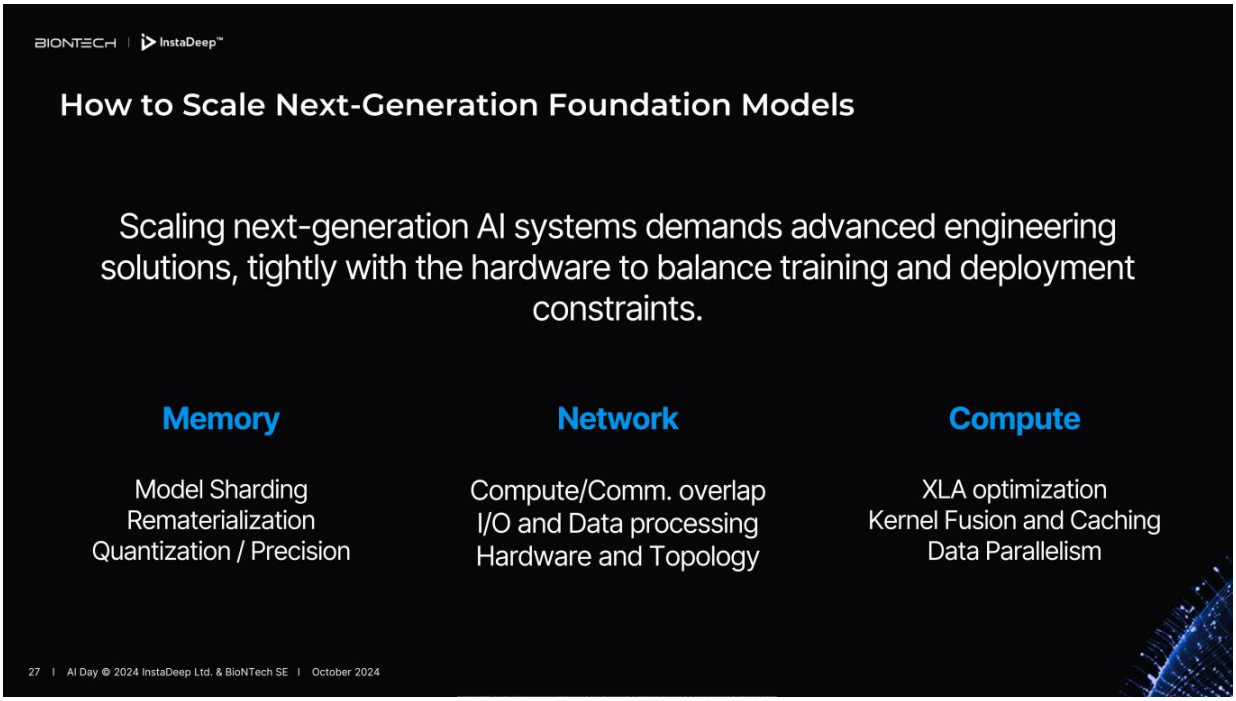
Scaling Laws

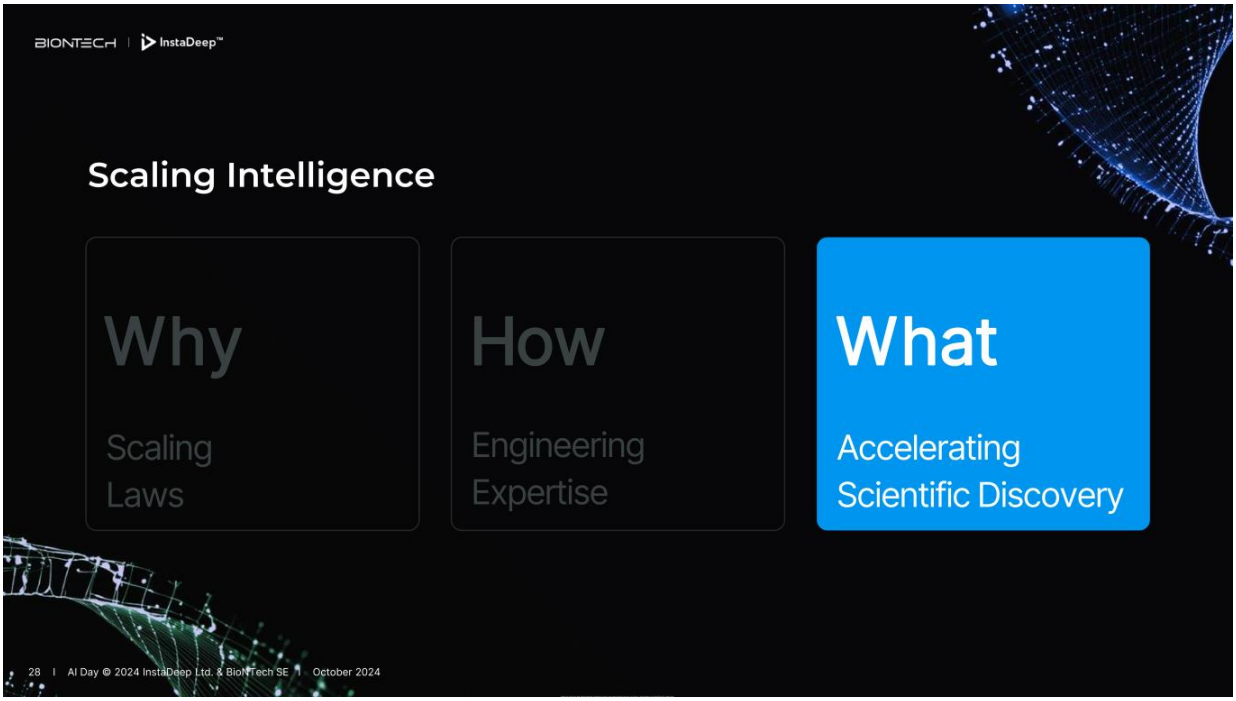
We can expect **"more intelligence" by scaling** existing algorithms.



Source: Achiam, Josh, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida et al. "Gpt-4 technical report." (2023)


BIONTECH | InstaDeep™
Scaling Intelligence
Why
Scaling Laws
How
Engineering Expertise
What
Accelerating Scientific Discovery
26 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
How to Scale Next-Generation Foundation Models
Scaling next-generation AI systems demands advanced engineering solutions, tightly with the hardware to balance training and deployment constraints.
Memory
Model Sharding
Rematerialization
Quantization / Precision
Network
Compute/Comm. overlap
I/O and Data processing
Hardware and Topology
Compute
XLA optimization
Kernel Fusion and Caching
Data Parallelism
27 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
Scaling Intelligence
Why
Scaling Laws
How
Engineering Expertise
What
Accelerating Scientific Discovery
28 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

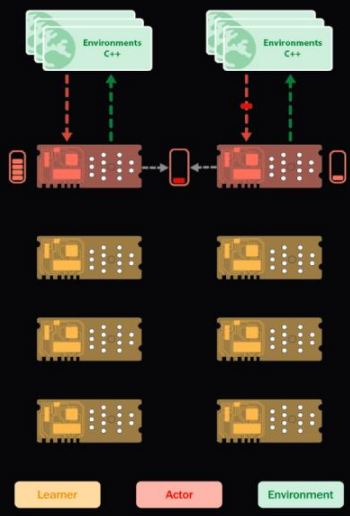
BIONTECH | InstaDeep™

#1 Accelerating Reinforcement Learning

Reinforcement Learning is the science of **learning from trials and errors.**
A simulation engine **turns computation into data.**

Scaling Reinforcement Learning

- Multiple threads keep the **hardware accelerators** active.
- Learner cores process experience, synchronizing updates using **JAX primitives.**
- The architecture can be **replicated across a large number of nodes** to form a supercomputing cluster.
- Leverage **the high-speed inter chip interconnects** between the nodes of the hardware accelerators.



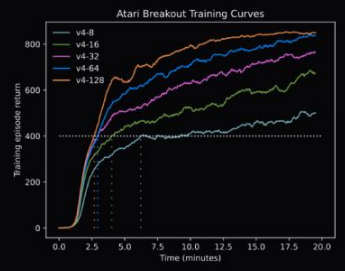
The Sebulba Architecture on 8x hardware accelerators[1]

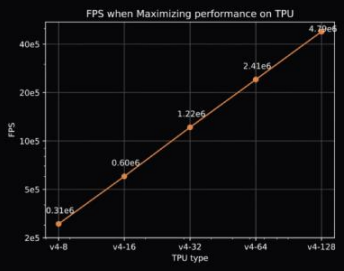
[1] "InstaDeep's scalable reinforcement learning on Cloud TPU", October 19, 2023, Google Cloud blog post
[2] Berner, C., Brockman, G., Chan, B., Cheung, V., Dębiak, P., Dennison, C., Farhi, D., Fischer, Q., Hashme, S., Hesse, C. and Józefowicz, R., 2019. Dota 2 with large scale deep reinforcement learning.
[3] Hessel, M., Kroiss, M., Clark, A., Kemaev, I., Quan, J., Keck, T., Viola, F. and van Hasselt, H., 2021. Podracer architectures for scalable reinforcement learning.

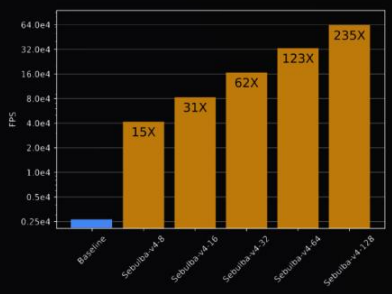
BIONTECH | InstaDeep™

#1 Accelerating Reinforcement Learning

Better: 50% improvement in performance as we scale the hardware and simulated data.
Cheaper: 13x cost reduction due to the more efficient use of the hardware.
Faster: 240x faster to train an RL agent up to convergence.







[1] "InstaDeep's scalable reinforcement learning on Cloud TPU", October 19, 2023, Google Cloud blog post
[2] Berner, C., Brockman, G., Chan, B., Cheung, V., Dębiak, P., Dennison, C., Farhi, D., Fischer, Q., Hashme, S., Hesse, C. and Józefowicz, R., 2019. Dota 2 with large scale deep reinforcement learning.
[3] Hessel, M., Kroiss, M., Clark, A., Kemaev, I., Quan, J., Keck, T., Viola, F. and van Hasselt, H., 2021. Podracer architectures for scalable reinforcement learning.

BIONTECH | InstaDeep™

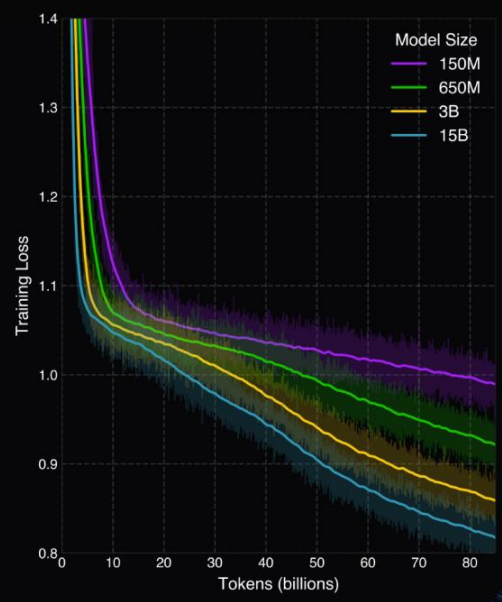
#2 Generative AI for Biology

Scaling the Next-Generation of our Generative AI Models

- in-house JAX-based software library
- best-in-class engineering for scaling LLMs
  e.g. Hybrid Parallelism, mixed-precision, rematerialization, etc.

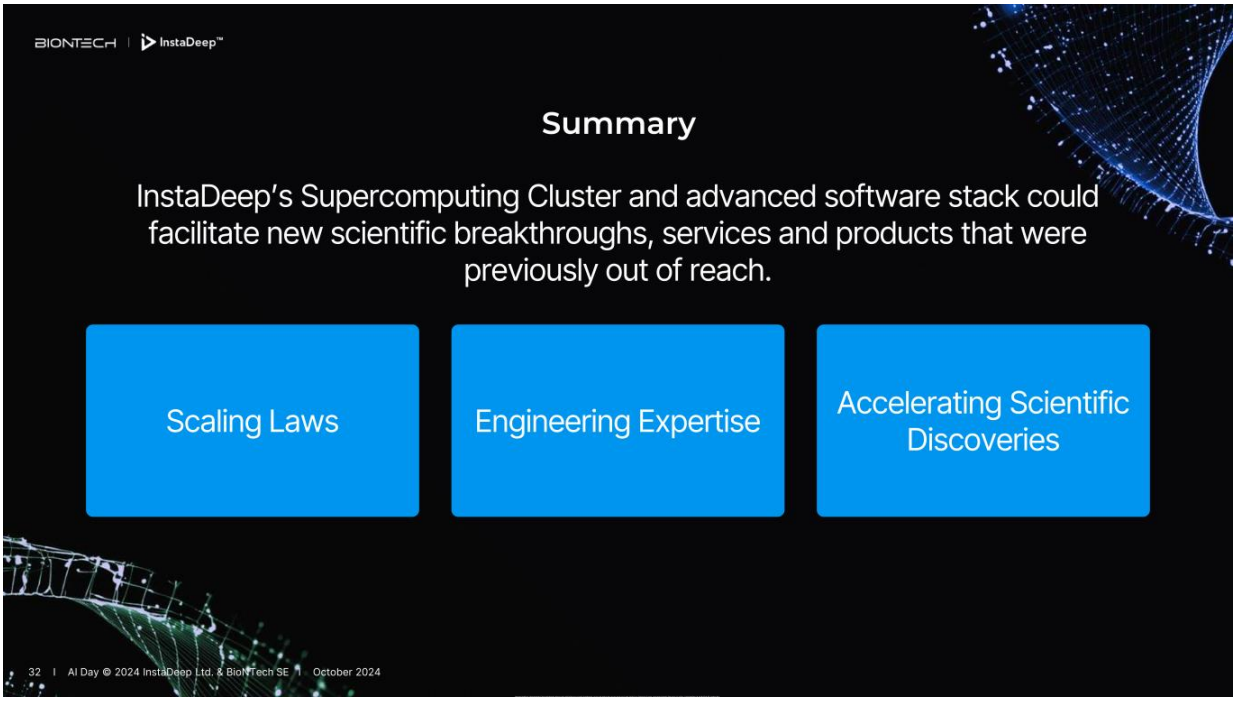
Results

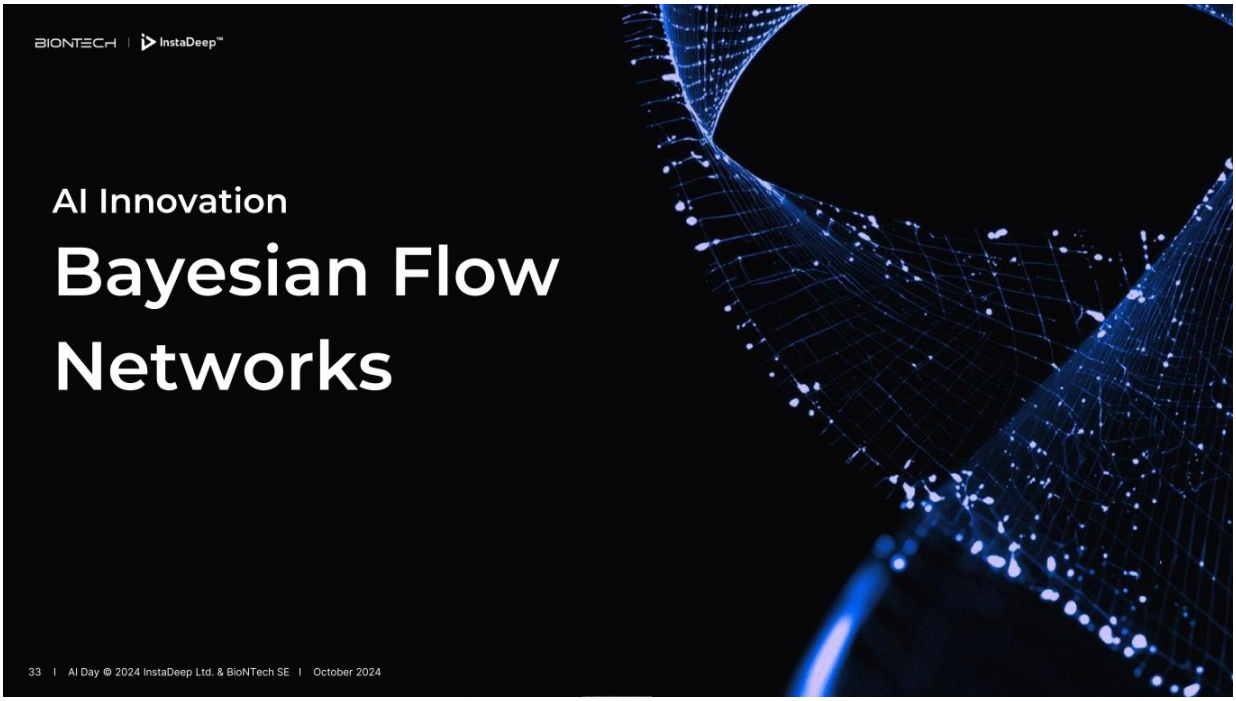
- Train multi-billion parameters models (+15B)
- Scaling Laws in Action
- Hardware efficiency on par with the latest Meta Llama 3.1[1]
  i.e. Model Flop Utilization of ~50% for our 15B model



Source: InstaDeep

[1] Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Yang, A., Fan, A. and Goyal, A., 2024. The llama 3 herd of models. arXiv preprint arXiv:2407.21783.


BIONTECH | InstaDeep™
Summary
InstaDeep's Supercomputing Cluster and advanced software stack could facilitate new scientific breakthroughs, services and products that were previously out of reach.
Scaling Laws
Engineering Expertise
Accelerating Scientific Discoveries
32 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
AI Innovation
Bayesian Flow Networks
33 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

BIONTECH | InstaDeep™

Generative AI

Sora: generative AI model for videos

OpenAI 2024

A stylish woman walks down a Tokyo street filled with warm glowing neon and animated city signage...

Source: SORA blog post by OpenAI: https://openai.com/index/sora/

BIONTECH | InstaDeep™

Joint Probability Distributions

For a generative model of face images, creating a new face means picking a sample from the joint probability distribution of all the pixels.

Q: Why is this so hard?

A: Because all the pixels are interrelated

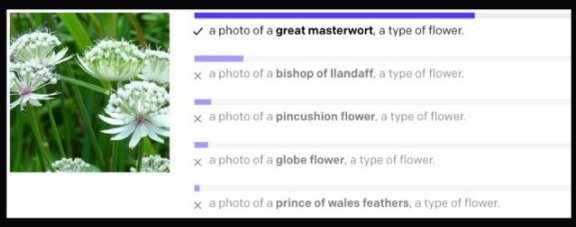
Generating Diverse High-Fidelity Images with **VQ-VAE-2** (Razavi et al. 2019)

BIONTECH | InstaDeep™

Steerable Generation

We want to control what we generate.

With a multimodal model (e.g. images and text) we can do this with conditional sampling: fix one modality, generate another.



Given an image, generate a caption (**CLIP** OpenAI 2021)

Given an prompt, generate an image (**DALL·E** OpenAI 2021)

BIONTECH | InstaDeep™

One Model, Many Tasks

Learning a joint distribution over many variables then choosing which to fix and which to generate gives us one model for many tasks

BIONTECH | InstaDeep™

But which Model?

Diffusion

Pros: Continuous data (especially images), inpainting, fast gradient-based sampling

Cons: Discrete data

Autoregression (GPT)

Sorry boss the dog ate my slides

Pros: Sequence data (especially text)

Cons: Unordered data, inpainting, slow sampling

Masked prediction (BERT)

Sorry ____ the ___ ate my _____

boss cat report

Pros: Discrete data, inpainting, representation learning

Cons: Continuous data, slow sampling

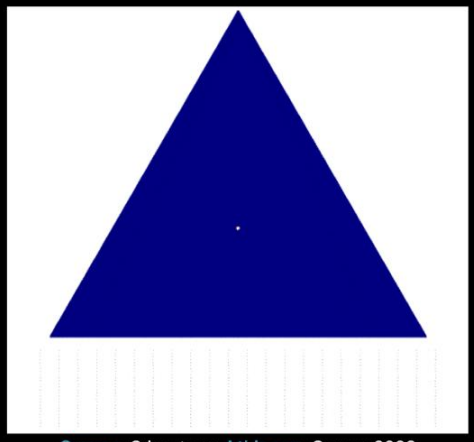
BIONTECH | InstaDeep™

Bayesian Flow Networks

BFNs are a new class of generative model that uses Bayesian inference to update beliefs about data.

Unlike diffusion models, they generate discrete data in a continuous way, allowing for gradient-based sampling

This makes BFNs well-suited for controllable generation across diverse data modalities.



Graves, Srivastava, Atkinson, Gomez 2023

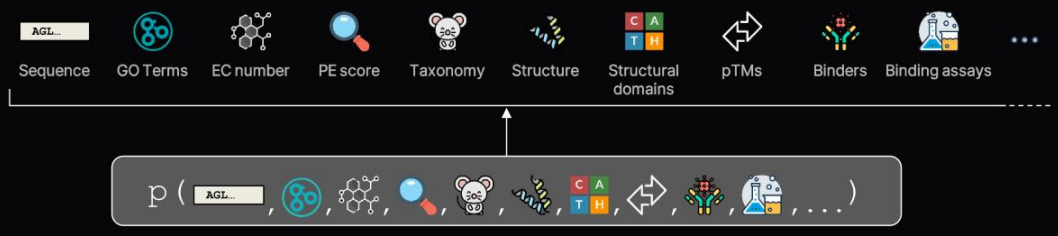
BIONTECH | InstaDeep™

Generative Modelling

A unifying framework to learn useful functions from data

(1) Learn to **model the joint distribution of *all* your data**, and (2) **conditionally sample for tasks of interest.**

Proteomics: Joint Modelling



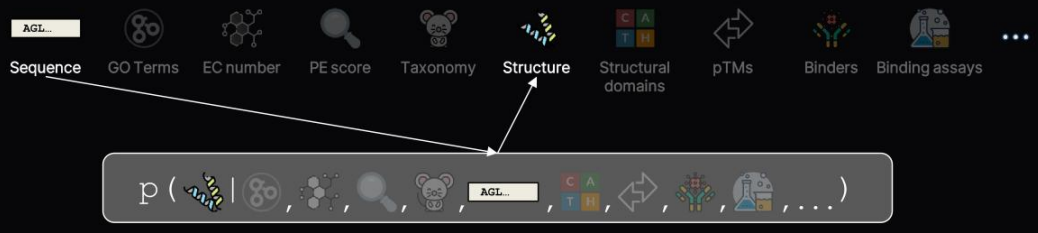
BIONTECH | InstaDeep™

Generative Modelling

A unifying framework to learn useful functions from data

(1) Learn to **model the joint distribution of *all* your data**, and (2) **conditionally sample for tasks of interest.**

Proteomics: Protein Folding

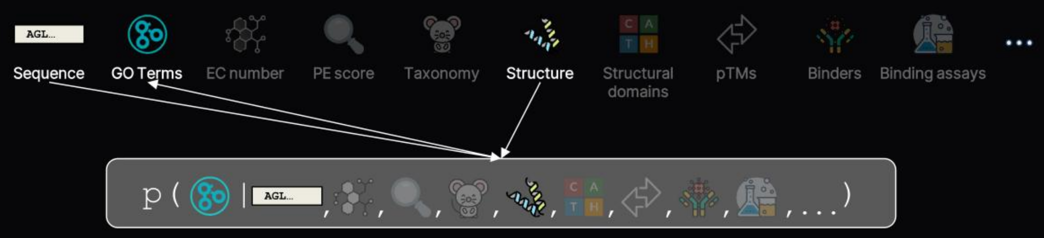


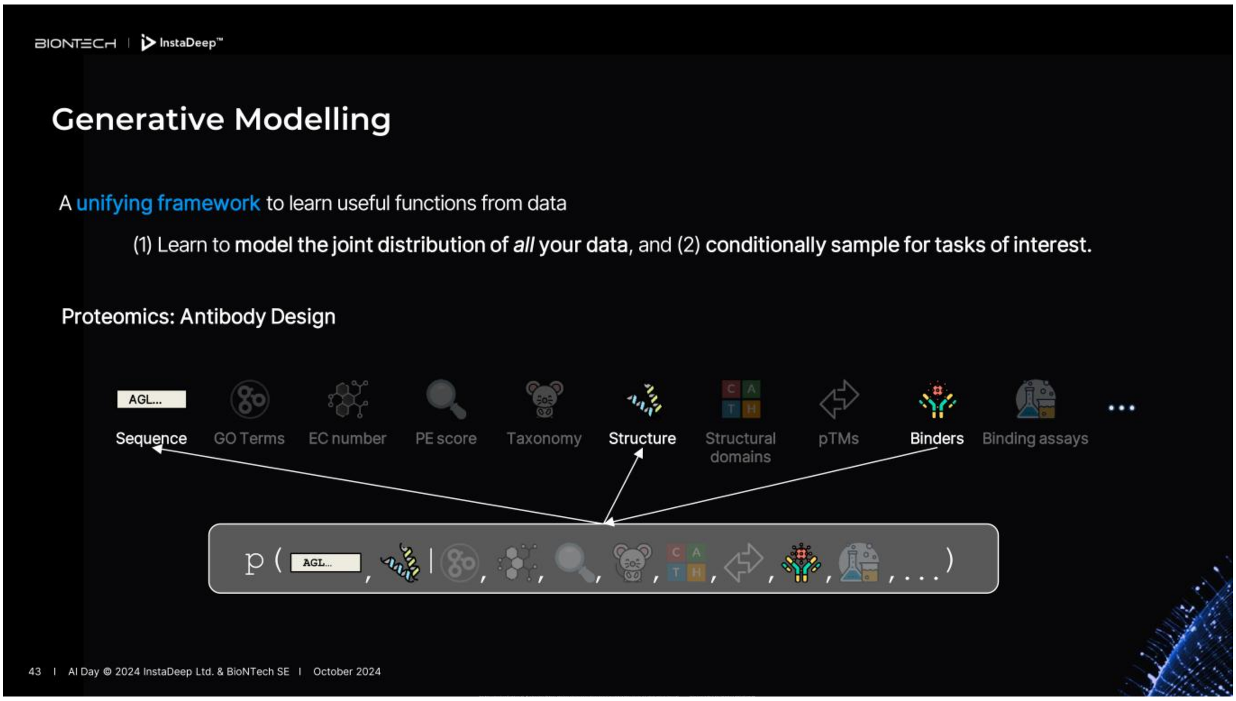
Generative Modelling

A unifying framework to learn useful functions from data

(1) Learn to **model the joint distribution of *all* your data**, and (2) **conditionally sample for tasks of interest.**

Proteomics: Function Prediction




BIONTECH | InstaDeep™
Generative Modelling
A unifying framework to learn useful functions from data
(1) Learn to model the joint distribution of all your data, and (2) conditionally sample for tasks of interest.
Proteomics: Antibody Design
AGL...
Sequence
GO Terms
EC number
PE score
Taxonomy
Structure
Structural domains
pTMs
Binders
Binding assays
...
p (AGL... , | , , , , , , , , ...)
43 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

BIONTECH | InstaDeep™

Generative Modelling

A **unifying framework** to learn useful functions from data

(1) Learn to **model the joint distribution of *all* your data**, and (2) **conditionally sample for tasks of interest.**

Proteomics: Sequence Generation

Sequence

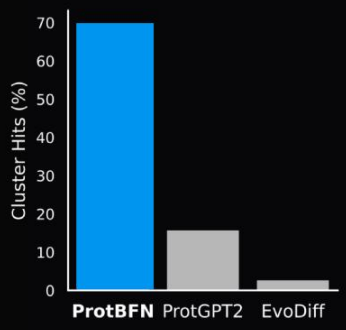
		De Novo	Conditional
Autoregression (GPT)	**R MPPR___...**	*Yes*	*Limited*
Masked prediction (BERT)	**RRS MPP___IV...**	*No*	*Yes*
"Discrete" diffusion	**P R I MMPRSSPV...**	*Limited applicability to discrete data*	

BIONTECH | InstaDeep™

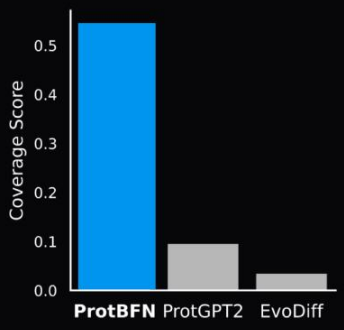
Natural, Diverse & Novel Protein Sequences

ProtBFN learns statistical and biochemical properties of natural proteins with high-fidelity.

More natural...[1]



...more diverse...[1]



...and highly novel.[2]

1. 10,000 generated sequences from each model are matched to clusterings from UniRef50. A hit is determined as a match with >50% sequence identity. Coverage score is the ratio of the number of unique clusters hit to the expected number if sequences were drawn i.i.d. from the models training distribution. ProtGPT2 (huggingface.co/nferruz/ProtGPT2) and EvoDiff (github.com/microsoft/evodiff) sequences are sampled using publicly available code and model weights provided by the authors.
2. Identity of ProtBFN generated sequences to the best matching protein sequence found in the models training data. Any identity < 100% is a novel sequence that the model has not seen before.

BIONTECH | InstaDeep™

Globular Structural Motifs With Novel Sequences

Predicted structures of generated sequences show natural, globally coherent and functionally diverse folds.

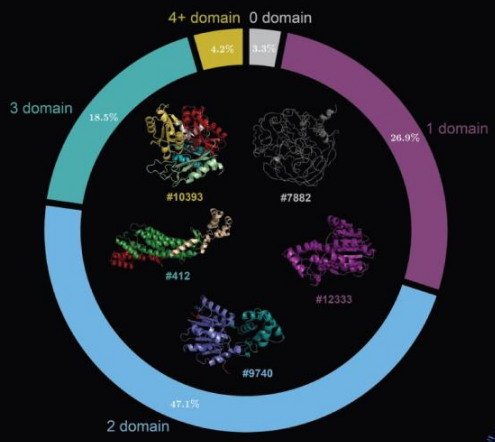
Structure largely determines function in nature.

Single and multi-domain proteins.

Globally coherent generations with inter-domain interactions.

Spans diversity of known structures and tree-of-life.

- Alpha Helical, Beta Sheet, Alpha-Beta and Irregular domains.
- Small and large domains.
- Transmembrane Proteins (porins and transporters) and Enzymes.
- Domains specific to Archaea, Bacteria, Eukarya (Plants, Humans).



BIONTECH | InstaDeep™

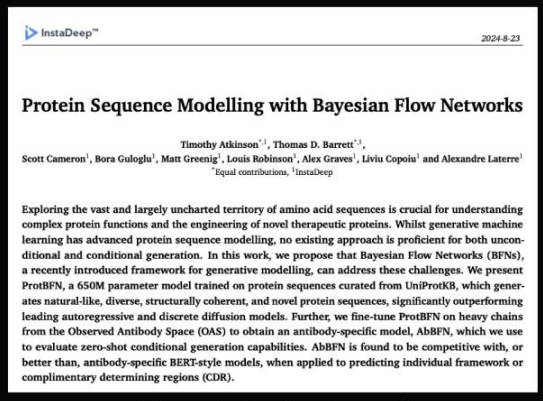
BFN for Protein Sequences

Outperforms or matches task-specific autoregressive, diffusion and BERT models.

Improved naturalness, diversity and novelty.

Uses zero-shot conditioning of model.

Patent application filed.



InstaDeep™ 2024-8-23

Protein Sequence Modelling with Bayesian Flow Networks

Timothy Atkinson[*,1], Thomas D. Barrett[*,1], Scott Cameron[1], Bora Guloglu[1], Matt Greenig[1], Louis Robinson[1], Alex Graves[1], Liviu Copoiu[1] and Alexandre Laterre[1]
[*]Equal contributions, [1]InstaDeep

Exploring the vast and largely uncharted territory of amino acid sequences is crucial for understanding complex protein functions and the engineering of novel therapeutic proteins. Whilst generative machine learning has advanced protein sequence modelling, no existing approach is proficient for both unconditional and conditional generation. In this work, we propose that Bayesian Flow Networks (BFNs), a recently introduced framework for generative modelling, can address these challenges. We present ProtBFN, a 650M parameter model trained on protein sequences curated from UniProtKB, which generates natural-like, diverse, structurally coherent, and novel protein sequences, significantly outperforming leading autoregressive and discrete diffusion models. Further, we fine-tune ProtBFN on heavy chains from the Observed Antibody Space (OAS) to obtain an antibody-specific model, AbBFN, which we use to evaluate zero-shot conditional generation capabilities. AbBFN is found to be competitive with, or better than, antibody-specific BERT-style models, when applied to predicting individual framework or complimentary determining regions (CDR).

Released only a few days ago![1]

1. Available at https://www.biorxiv.org/content/10.1101/2024.09.24.614734v1

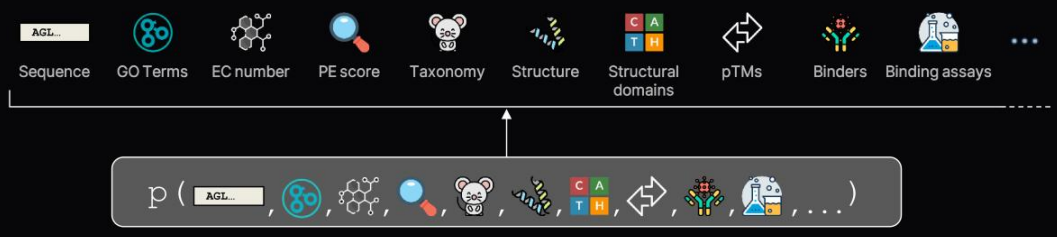
BIONTECH | InstaDeep™

Going Beyond Sequence-Only Models

Our goal is to model everything: building foundational models of the joint distribution of heterogeneous scientific data.

Performance across multiple data types and sources.

Flexibility in the hands of scientists with task-specific inference.



BIONTECH | InstaDeep™

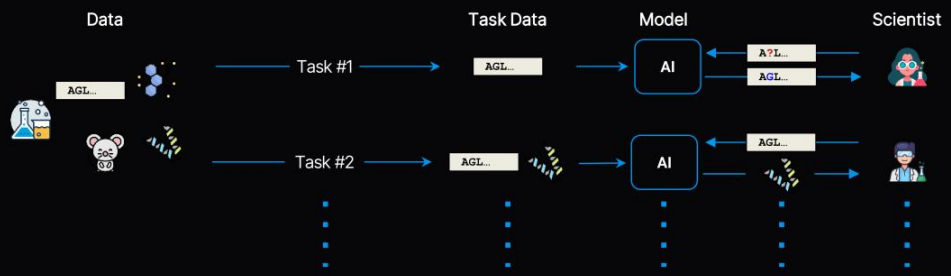
Going Beyond Sequence-Only Models

Our goal is to model everything: building foundational models of the joint distribution of heterogeneous scientific data.

Performance across multiple data types and sources.

Flexibility in the hands of scientists with task-specific inference.

Conventional ML



BIONTECH | InstaDeep™

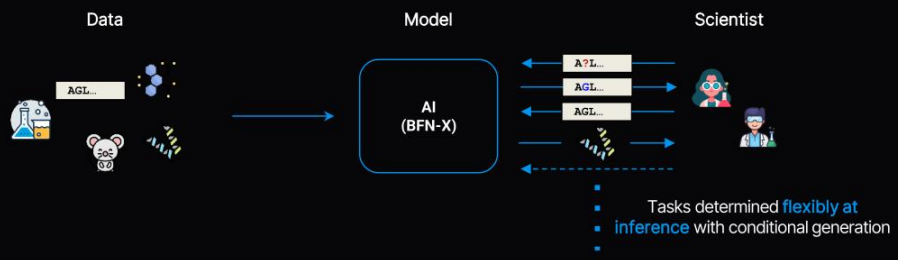
Going Beyond Sequence-Only Models

Our goal is to model everything: building foundational models of the joint distribution of heterogeneous scientific data.

Performance across multiple data types and sources.

Flexibility in the hands of scientists with task-specific inference.

Our vision



Tasks determined **flexibly at inference** with conditional generation

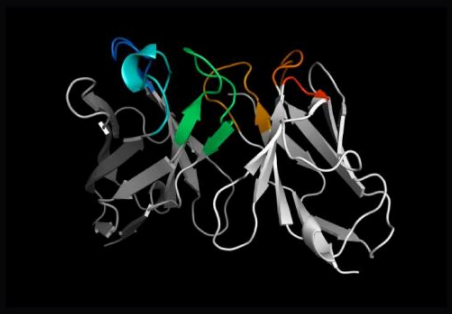
BIONTECH | InstaDeep™

Introducing AbBFN-X

First look at our **multimodal model** for antibodies.
36 different attributes jointly modelled: **sequence**, **genetic**, **biophysical**

Empowers scientists with **tunable generation**, being **highly flexible** across many tasks.

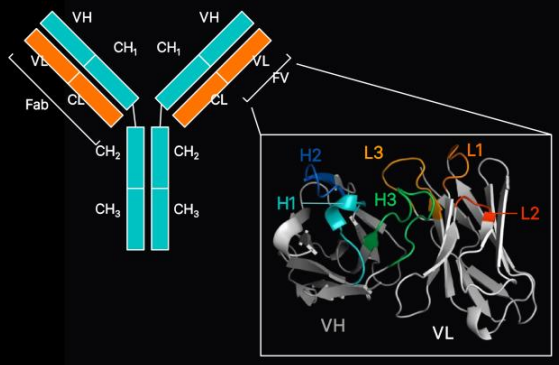
Today's use cases go **beyond standard AI-enhanced antibody design** workflows.



Source: Generated Image

BIONTECH | InstaDeep™

AbBFN-X

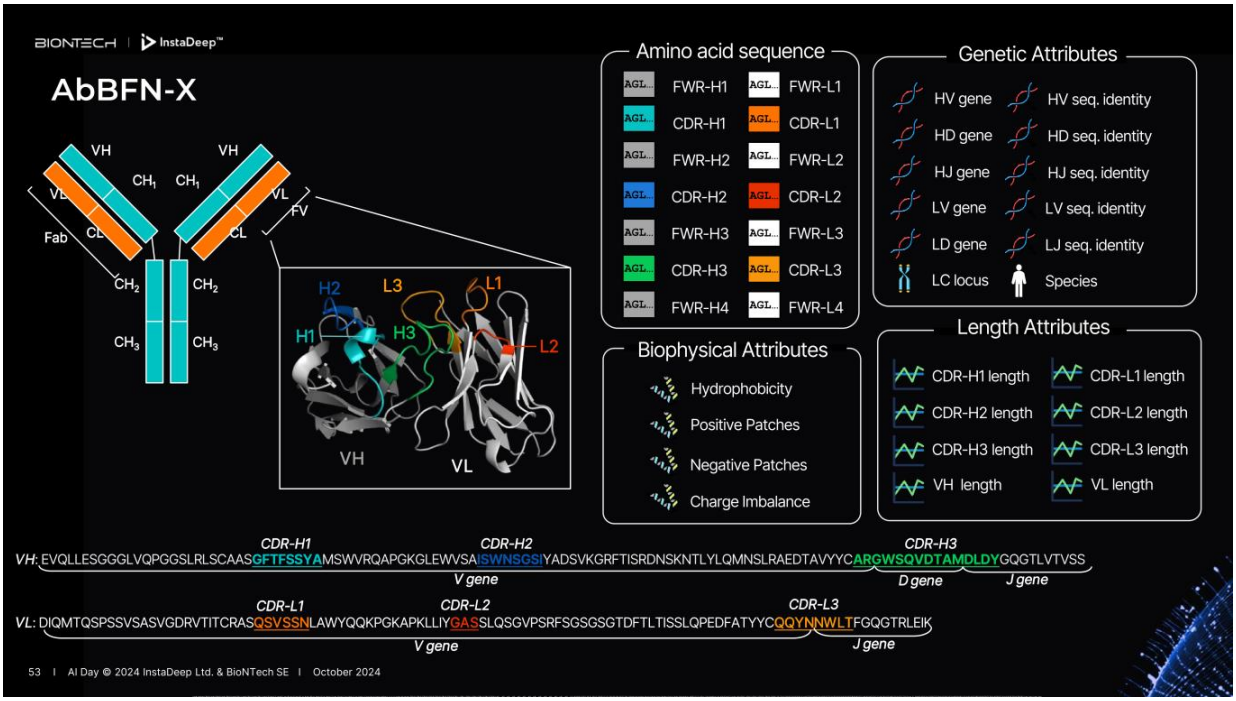


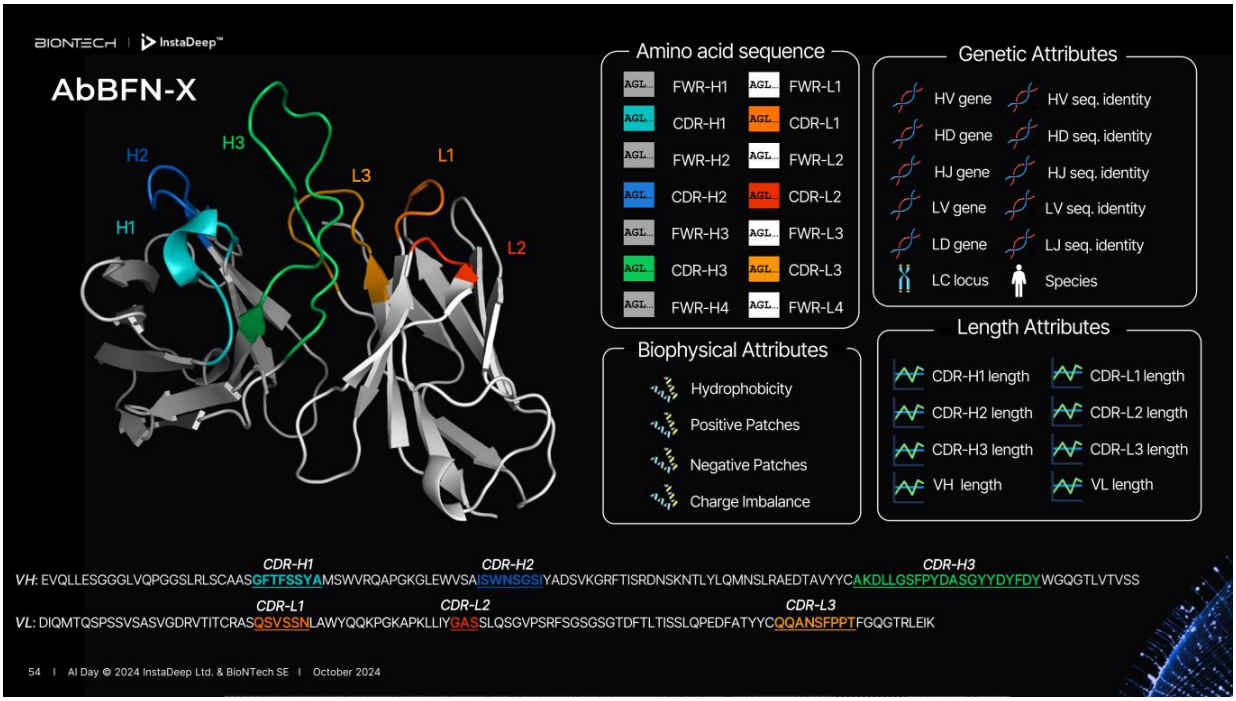
VH: EVQLLESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVRQAPGKGLEWVSAISWNSGSIYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYCARGWSQVDTAMDLDYGQGTLVTVSS

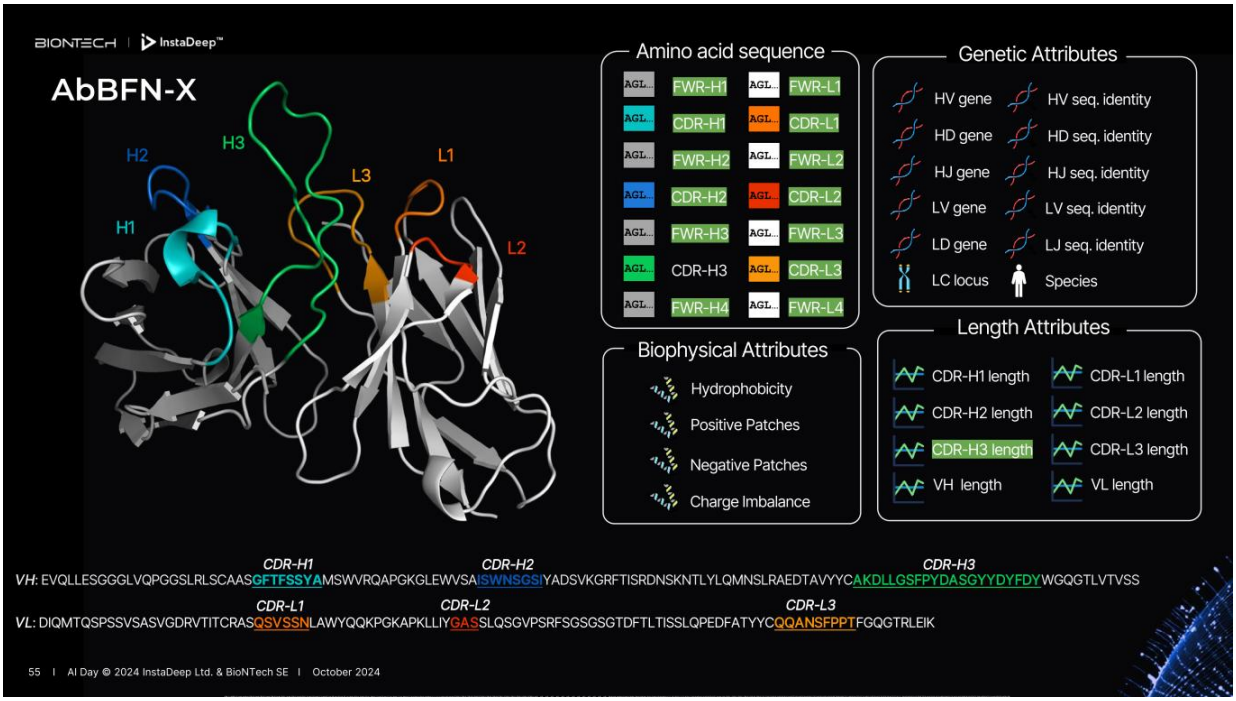
CDR-H1 · *CDR-H2* · *CDR-H3* · *V gene* · *D gene* · *J gene*

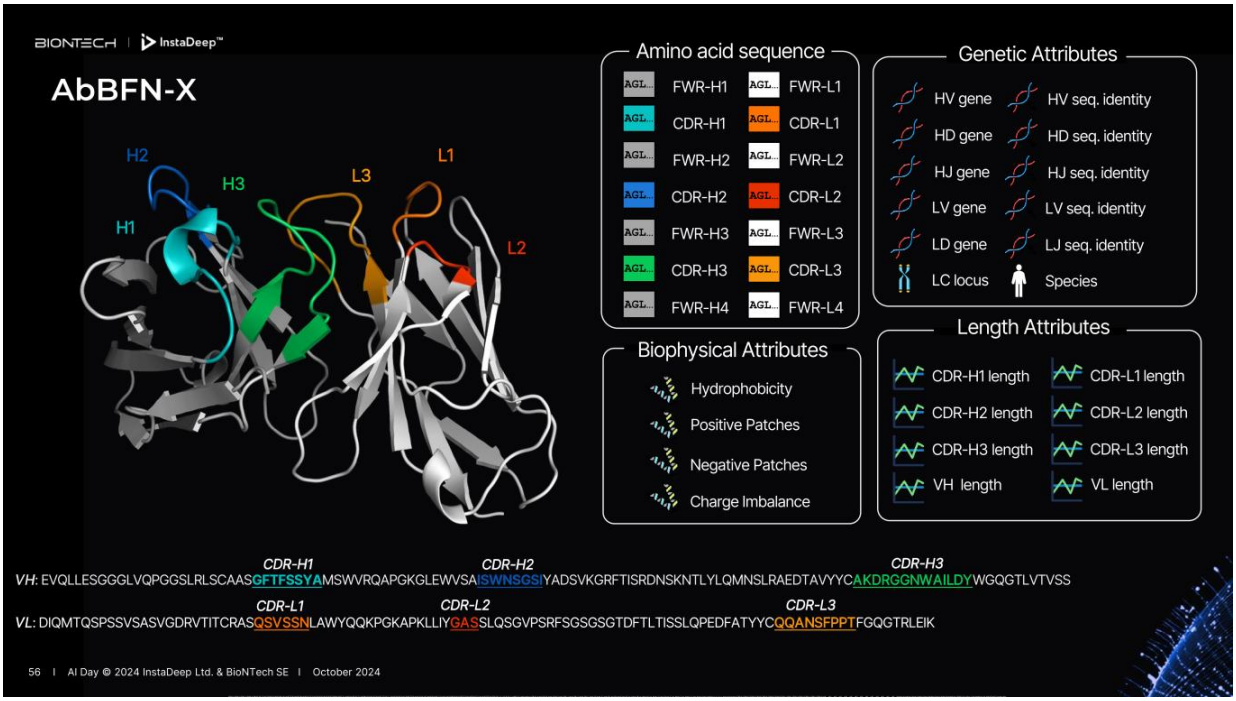
VL: DIQMTQSPSSVSASVGDRVTITCRASQSVSSNLAWYQQKPGKAPKLLIYGASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQYNNWLTFGQGTRLEIK

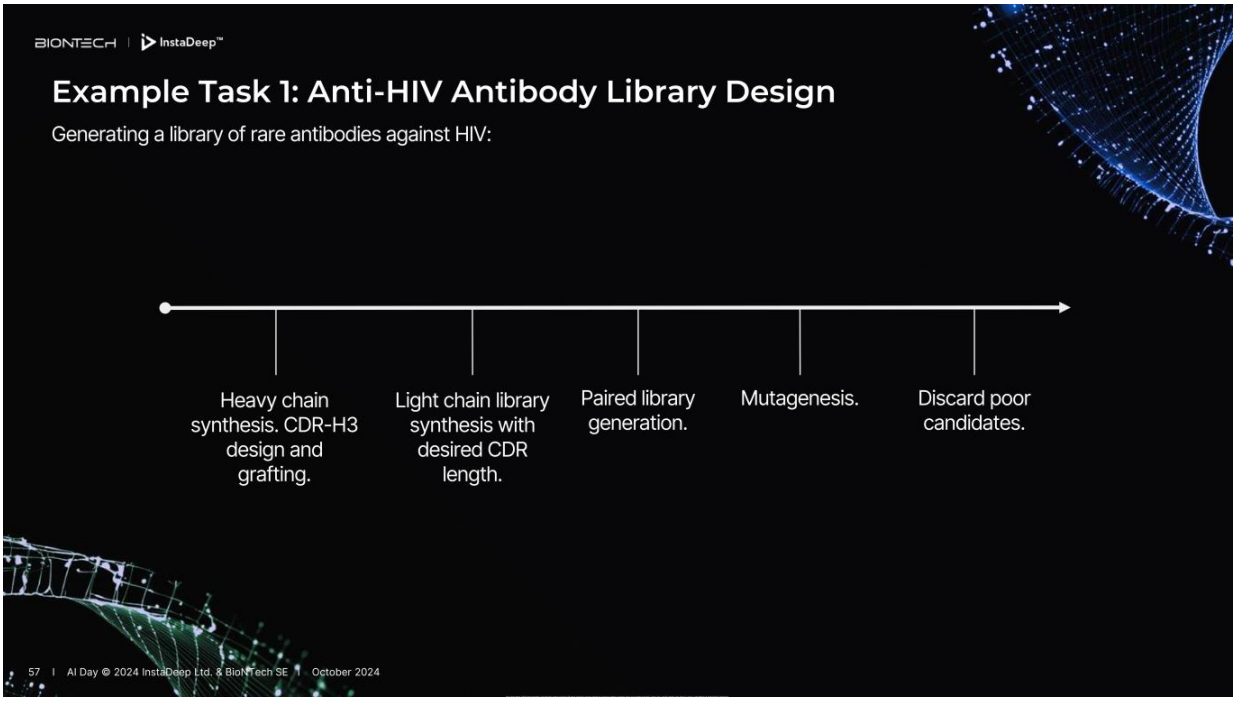
CDR-L1 · *CDR-L2* · *CDR-L3* · *V gene* · *J gene*

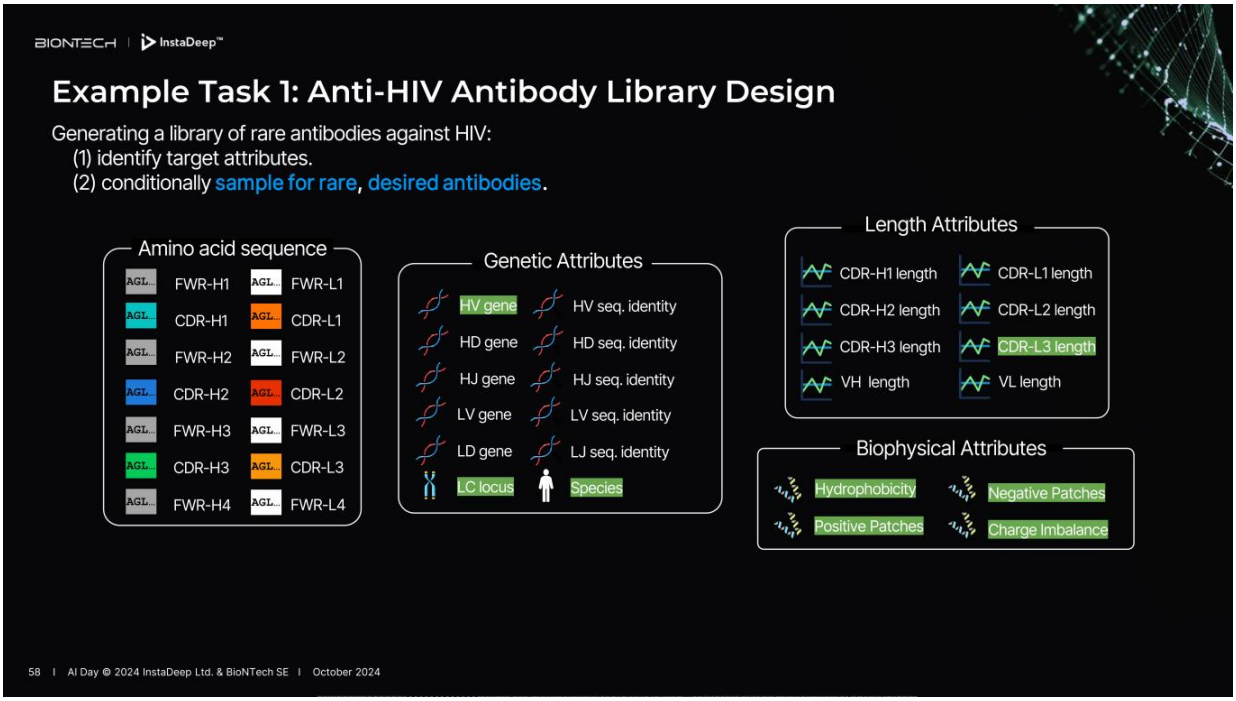

BIONTECH | InstaDeep™
AbBFN-X
VH
VH
CH1
CH1
VL
VL
FV
Fab
CL
CL
CH2
CH2
CH3
CH3
H2
L3
L1
H1
H3
L2
VH
VL
Amino acid sequence
FWR-H1
FWR-L1
CDR-H1
CDR-L1
FWR-H2
FWR-L2
CDR-H2
CDR-L2
FWR-H3
FWR-L3
CDR-H3
CDR-L3
FWR-H4
FWR-L4
Genetic Attributes
HV gene
HV seq. identity
HD gene
HD seq. identity
HJ gene
HJ seq. identity
LV gene
LV seq. identity
LD gene
LJ seq. identity
LC locus
Species
Biophysical Attributes
Hydrophobicity
Positive Patches
Negative Patches
Charge Imbalance
Length Attributes
CDR-H1 length
CDR-L1 length
CDR-H2 length
CDR-L2 length
CDR-H3 length
CDR-L3 length
VH length
VL length
CDR-H1
CDR-H2
CDR-H3
VH: EVQLLESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVRQAPGKGLEWVSAISWNSGSIYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYCARGWSQVDTAMDLDYGQGTLVTVSS
V gene
D gene
J gene
CDR-L1
CDR-L2
CDR-L3
VL: DIQMTQSPSSVSASVGDRVTITCRASQSVSSNLAWYQQKPGKAPKLLIYGASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQYNNWLTFGQGTRLEIK
V gene
J gene
53 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
AbBFN-X
H1
H2
H3
L1
L2
L3
Amino acid sequence
FWR-H1
CDR-H1
FWR-H2
CDR-H2
FWR-H3
CDR-H3
FWR-H4
FWR-L1
CDR-L1
FWR-L2
CDR-L2
FWR-L3
CDR-L3
FWR-L4
Genetic Attributes
HV gene
HD gene
HJ gene
LV gene
LD gene
LC locus
HV seq. identity
HD seq. identity
HJ seq. identity
LV seq. identity
LJ seq. identity
Species
Biophysical Attributes
Hydrophobicity
Positive Patches
Negative Patches
Charge Imbalance
Length Attributes
CDR-H1 length
CDR-H2 length
CDR-H3 length
VH length
CDR-L1 length
CDR-L2 length
CDR-L3 length
VL length
CDR-H1
CDR-H2
CDR-H3
VH: EVQLLESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVRQAPGKGLEWVSAISWNSGSIYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYCAKDLLGSFPYDASGYYDYFDYWGQGTLVTVSS
CDR-L1
CDR-L2
CDR-L3
VL: DIQMTQSPSSVSASVGDRVTITCRASQSVSSNLAWYQQKPGKAPKLLIYGASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQANSFPPTFGQGTRLEIK
54 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
AbBFN-X
H1
H2
H3
L1
L2
L3
Amino acid sequence
FWR-H1
CDR-H1
FWR-H2
CDR-H2
FWR-H3
CDR-H3
FWR-H4
FWR-L1
CDR-L1
FWR-L2
CDR-L2
FWR-L3
CDR-L3
FWR-L4
Genetic Attributes
HV gene
HD gene
HJ gene
LV gene
LD gene
LC locus
HV seq. identity
HD seq. identity
HJ seq. identity
LV seq. identity
LJ seq. identity
Species
Biophysical Attributes
Hydrophobicity
Positive Patches
Negative Patches
Charge Imbalance
Length Attributes
CDR-H1 length
CDR-H2 length
CDR-H3 length
VH length
CDR-L1 length
CDR-L2 length
CDR-L3 length
VL length
CDR-H1
CDR-H2
CDR-H3
VH: EVQLLESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVRQAPGKGLEWVSAISWNSGSIYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYCAKDLLGSFPYDASGYYDYFDYWGQGTLVTVSS
CDR-L1
CDR-L2
CDR-L3
VL: DIQMTQSPSSVSASVGDRVTITCRASQSVSSNLAWYQQKPGKAPKLLIYGASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQANSFPPTFGQGTRLEIK
55 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
AbBFN-X
H2
H3
H1
L3
L1
L2
Amino acid sequence
FWR-H1
FWR-L1
CDR-H1
CDR-L1
FWR-H2
FWR-L2
CDR-H2
CDR-L2
FWR-H3
FWR-L3
CDR-H3
CDR-L3
FWR-H4
FWR-L4
Genetic Attributes
HV gene
HV seq. identity
HD gene
HD seq. identity
HJ gene
HJ seq. identity
LV gene
LV seq. identity
LD gene
LJ seq. identity
LC locus
Species
Biophysical Attributes
Hydrophobicity
Positive Patches
Negative Patches
Charge Imbalance
Length Attributes
CDR-H1 length
CDR-L1 length
CDR-H2 length
CDR-L2 length
CDR-H3 length
CDR-L3 length
VH length
VL length
CDR-H1
CDR-H2
CDR-H3
VH: EVQLLESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVRQAPGKGLEWVSAISWNSGSIYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYCAKDRGGNWAILDYWGQGTLVTVSS
CDR-L1
CDR-L2
CDR-L3
VL: DIQMTQSPSSVSASVGDRVTITCRASQSVSSNLAWYQQKPGKAPKLLIYGASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQANSFPPTFGQGTRLEIK
56 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

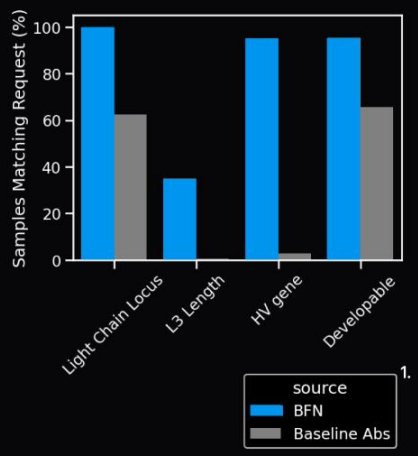

BIONTECH | InstaDeep™
Example Task 1: Anti-HIV Antibody Library Design
Generating a library of rare antibodies against HIV:
Heavy chain synthesis. CDR-H3 design and grafting.
Light chain library synthesis with desired CDR length.
Paired library generation.
Mutagenesis.
Discard poor candidates.
57 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
Example Task 1: Anti-HIV Antibody Library Design
Generating a library of rare antibodies against HIV:
(1) identify target attributes.
(2) conditionally sample for rare, desired antibodies.
Amino acid sequence
FWR-H1
FWR-L1
CDR-H1
CDR-L1
FWR-H2
FWR-L2
CDR-H2
CDR-L2
FWR-H3
FWR-L3
CDR-H3
CDR-L3
FWR-H4
FWR-L4
Genetic Attributes
HV gene
HV seq. identity
HD gene
HD seq. identity
HJ gene
HJ seq. identity
LV gene
LV seq. identity
LD gene
LJ seq. identity
LC locus
Species
Length Attributes
CDR-H1 length
CDR-L1 length
CDR-H2 length
CDR-L2 length
CDR-H3 length
CDR-L3 length
VH length
VL length
Biophysical Attributes
Hydrophobicity
Negative Patches
Positive Patches
Charge Imbalance
58 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

BIONTECH | InstaDeep™

Example Task 1: Anti-HIV Antibody Library Design

Generating a library of rare antibodies against HIV:
(1) identify target attributes.
(2) conditionally sample for rare, desired antibodies.



1.

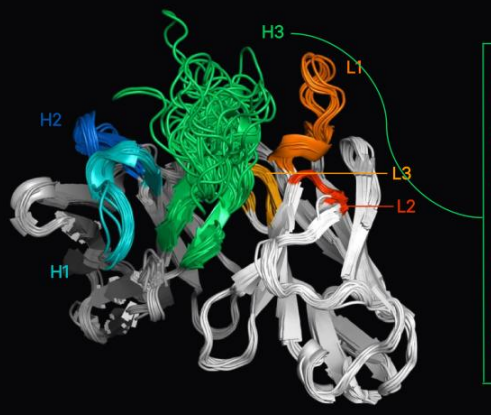
AbBFN-X antibodies are 5600x more likely[2] to have all desired characteristics.

1. "Baseline Abs" refers to data sets of natural antibodies (Olsen *et al*, 2021, Prot. Sci.), "BFN" refers to samples generated by AbBFN-X.
2. compared to the rate of finding antibodies with the correct characteristics in data sets of natural antibodies (Olsen *et al*, 2021, Prot. Sci.)

BIONTECH | InstaDeep™

Example Task 1: Anti-HIV Antibody Library Design

Generating a library of rare antibodies against HIV:
(1) identify target attributes.
(2) conditionally sample for rare, desired antibodies.



ARDEIYFLEWLISY
AKVRLGELPYEAFDI
ARGVRVQ SYNWFDP
ASGEYFFDTSSYPN
ARSSFVYPKSGYDFYFDY
ARDIAVDPESTAYFDY
AKGFSYGDGWADY
VRLRVGVLPGAFDI
ARDGGHYSH
ASGSGDSRYAQPLWFTTAFDI
ATSLNYGVIISD
ASGKMAVAYYFDY
AREGMDASMYYFDY
ARDMGYHDGALVFDN
⋮

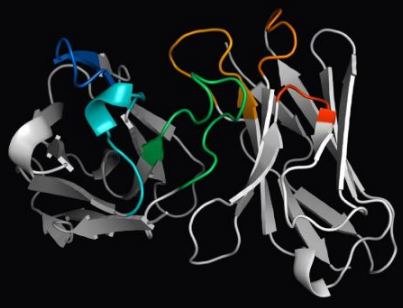
Unique CDRs[1]	100%
Unique CDR-H3s	100%
Unique CDR-L3s	99%
Unique non H/L3 CDRs	52%

1. 128 samples were generated, uniqueness assessed by considering all relevant regions at once, excluding framework regions.

BIONTECH | InstaDeep™

Example Task 2: Heavy-Light Pairing

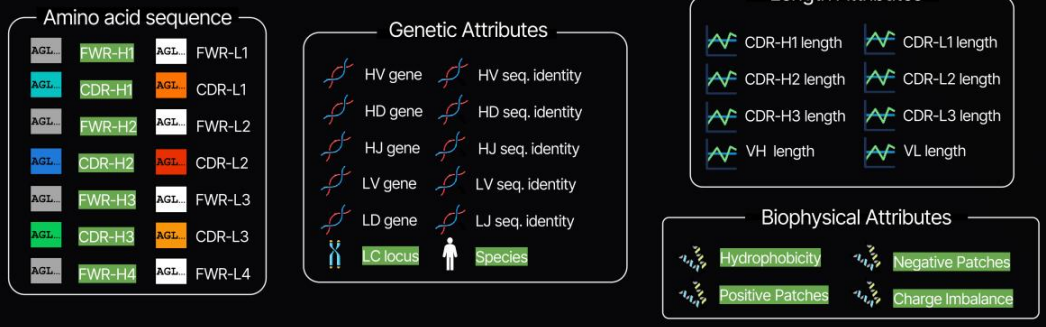
Generating a set of **developable** light chains that will **pair** with a heavy chain:
(1) **Condition** on desired properties and heavy sequence.
(2) **Sample** for **stable**, **diverse** sequences.



BIONTECH | InstaDeep™

Example Task 2: Heavy-Light Pairing

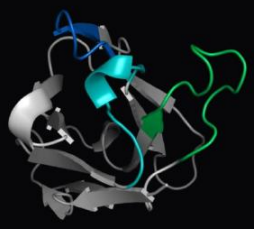
Generating a set of **developable** light chains that will **pair** with a heavy chain:
(1) **Condition** on desired properties and heavy sequence.
(2) **Sample** for **stable**, **diverse** sequences.



BIONTECH | InstaDeep™

Example Task 2: Heavy-Light Pairing

Generating a set of **developable** light chains that will **pair** with a heavy chain:
(1) **Condition** on desired properties and heavy sequence.
(2) **Sample** for **stable**, **diverse** sequences.

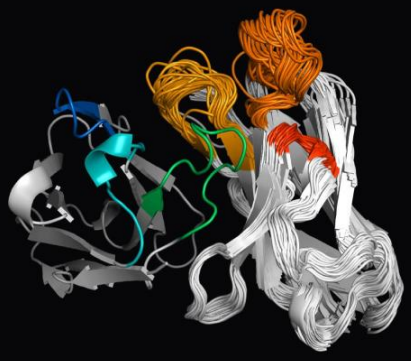


Conditioning on the required **heavy chain**

BIONTECH | InstaDeep™

Example Task 2: Heavy-Light Pairing

Generating a set of **developable** light chains that will **pair** with a heavy chain:
(1) **Condition** on desired properties and heavy sequence.
(2) **Sample** for **stable**, **diverse** sequences.



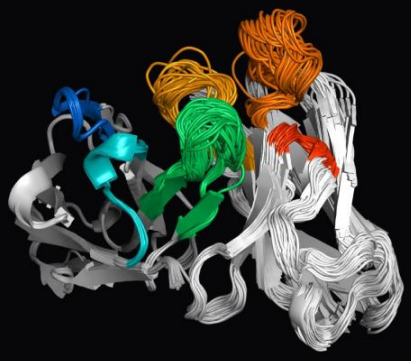
Conditioning on the required **heavy chain**

results in **diverse light chains** in length and structure

BIONTECH | InstaDeep™

Example Task 2: Heavy-Light Pairing

Generating a set of **developable** light chains that will **pair** with a heavy chain:
(1) **Condition** on desired properties and heavy sequence.
(2) **Sample** for **stable**, **diverse** sequences.



Conditioning on the required **heavy chain**

results in **diverse light chains** in length and structure

while respecting the **requested sequence** and generating **stable pairs**.

BIONTECH | InstaDeep™

Generative AI for Proteomics

InstaDeep is developing **next-generation GenAI models** across the whole stack, from **fundamental ML research**, to modelling scientific data and **enabling new capabilities** for scientists.

Bayesian Flow Networks
Unified modelling of multi-modal data

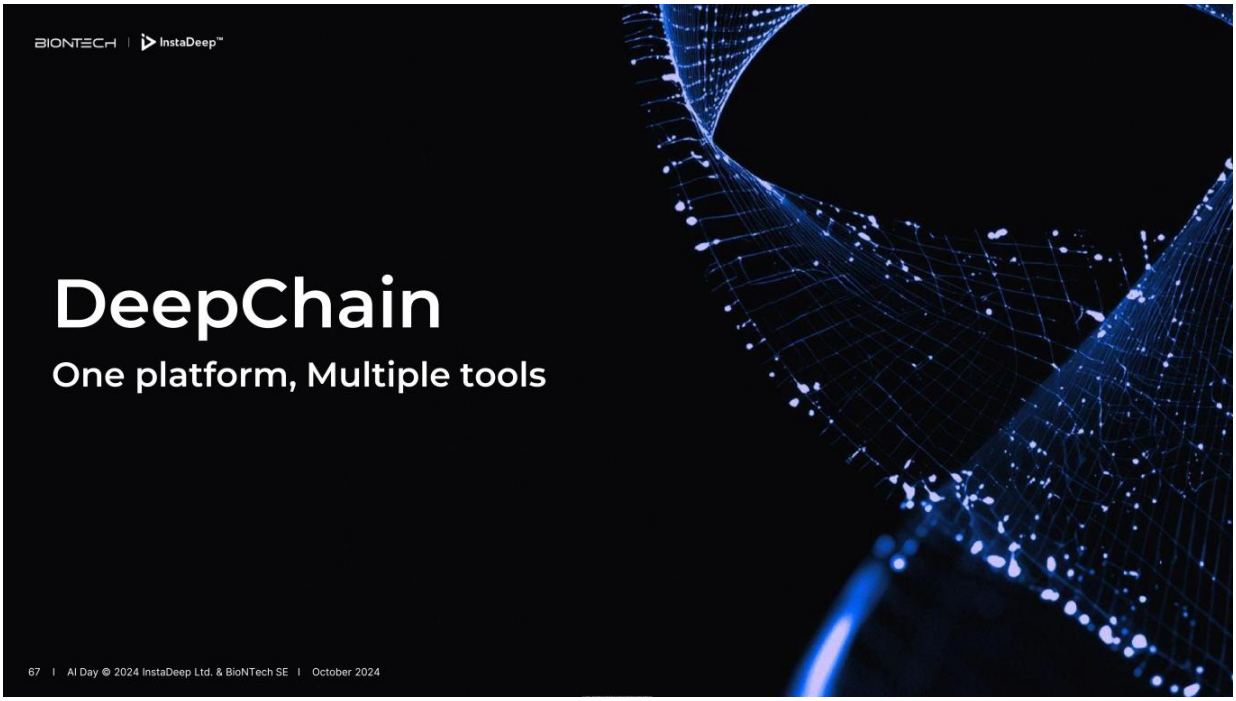
- **All modalities** as first class citizens
- Joint learning of **heterogeneous data**
- Task-specific **conditional generation**

Protein Sequence Modelling
Published demonstrations

- **Leading performance** across tasks
- Diverse & novel ***de novo*** **generation**
- **Zero-shot** inpainting at inference

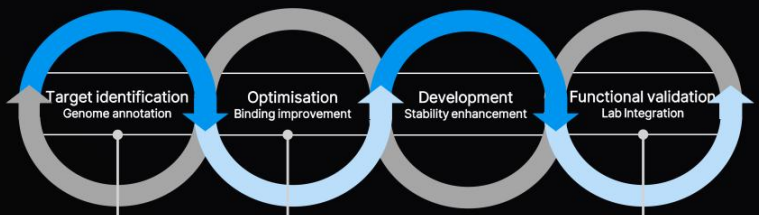
BFN-X
In-development foundation models

- **Sequence**, **genetic** and **biophysical**
- Learns **rational antibody principles**
- Enables **diverse suite of tasks**


BIONTECH | InstaDeep™
DeepChain
One platform, Multiple tools
67 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

BIONTECH | InstaDeep™

By combining State-of-the-art Science with Engineering, Our AI Tools aim to Accelerate the R&D pipeline



Nucleotide Transformers[2,3]

- Splicing prediction
- In silico design of regulatory sequences

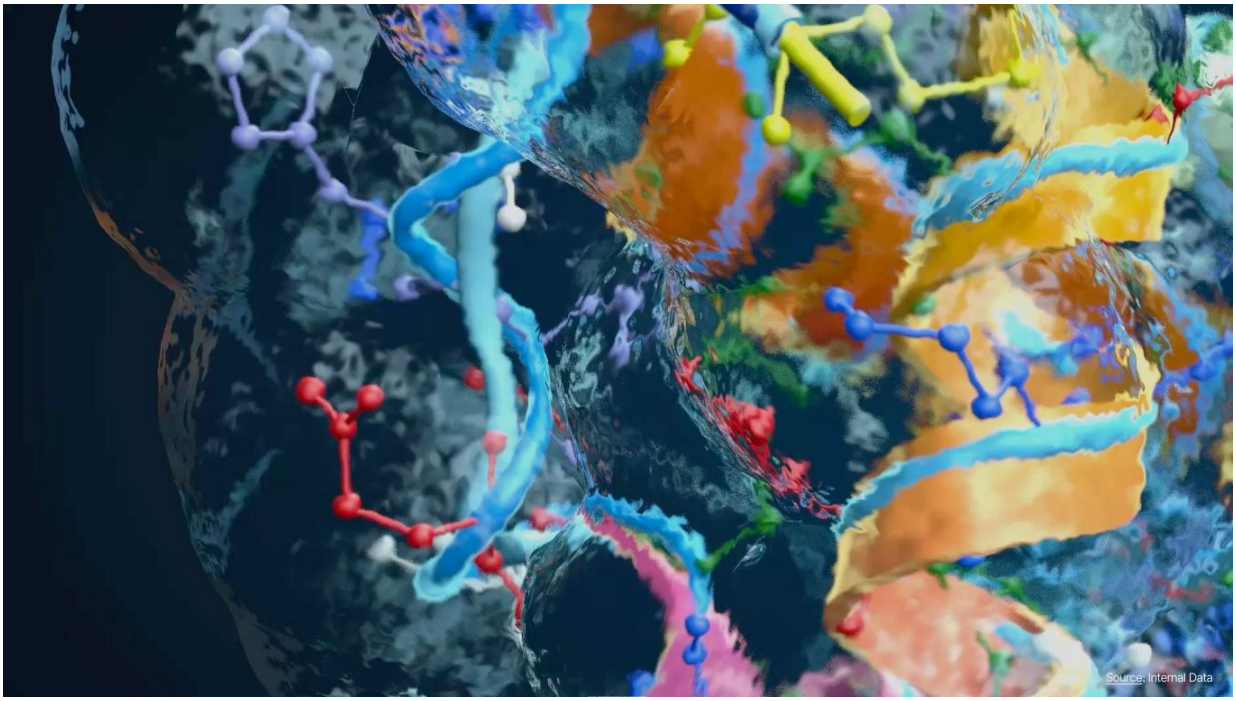
Bayesian Flow Networks[1]

- De novo antibody design
- Affinity enhancement

Assistants

- Form hypotheses, design experiments - Call tools to analyse results collaborating with a human scientist

[1] Alice Sends Amino Acids to Bob: Protein Sequence Modelling with Bayesian Flow Networks, Barrett et al., Under Review (2024).
[2] The Nucleotide Transformer: Building and Evaluating Robust Foundation Models for Human Genomics, Dalla-Torre et al., Under Review (2023).
[3] SegmentNT: annotating the genome at single-nucleotide resolution with DNA foundation models, Almeida et al., Under Review (2024).


Source: Internal Data

BIONTECH | InstaDeep™

We Are Releasing our Flagship Models on DeepChain

ProtBFN & AbBFN[1]

State-of-the-art generative protein models

✓ Generates natural-like, diverse, structurally coherent, and novel protein sequences

✓ Outperforms leading autoregressive and discrete diffusion models

✓ Enables flexible conditional generation in a zero-shot manner

Nucleotide Transformer[2] & SegmentNT[3]

Our foundation models for DNA

✓ Single nucleotide resolution

✓ Up to 50kb context length without performance drop

✓ Generalizes across species

[1] Alice Sends Amino Acids to Bob: Protein Sequence Modelling with Bayesian Flow Networks, Barrett et al., Under Review (2024).
[2] The Nucleotide Transformer: Building and Evaluating Robust Foundation Models for Human Genomics, Dalla-Torre et al., Under Review (2023).
[3] SegmentNT: annotating the genome at single-nucleotide resolution with DNA foundation models, Almeida et al., Under Review (2024).

BIONTECH | InstaDeep™

We Are Releasing our Flagship Models on DeepChain

ProtBFN & AbBFN[1]

State-of-the-art generative protein models

✓ Generates natural-like, diverse, structurally coherent, and novel protein sequences

✓ Outperforms leading autoregressive and discrete diffusion models

✓ Enables flexible conditional generation in a zero-shot manner

Nucleotide Transformer[2] & SegmentNT[3]

Our foundation models for DNA

✓ Single nucleotide resolution

✓ Up to 50kb context length without performance drop

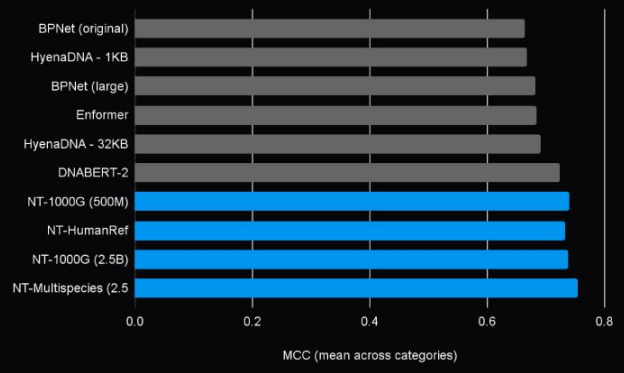
✓ Generalizes across species

[1] Alice Sends Amino Acids to Bob: Protein Sequence Modelling with Bayesian Flow Networks, Barrett et al., Under Review (2024).
[2] The Nucleotide Transformer: Building and Evaluating Robust Foundation Models for Human Genomics, Dalla-Torre et al., Under Review (2023).
[3] SegmentNT: annotating the genome at single-nucleotide resolution with DNA foundation models, Almeida et al., Under Review (2024).

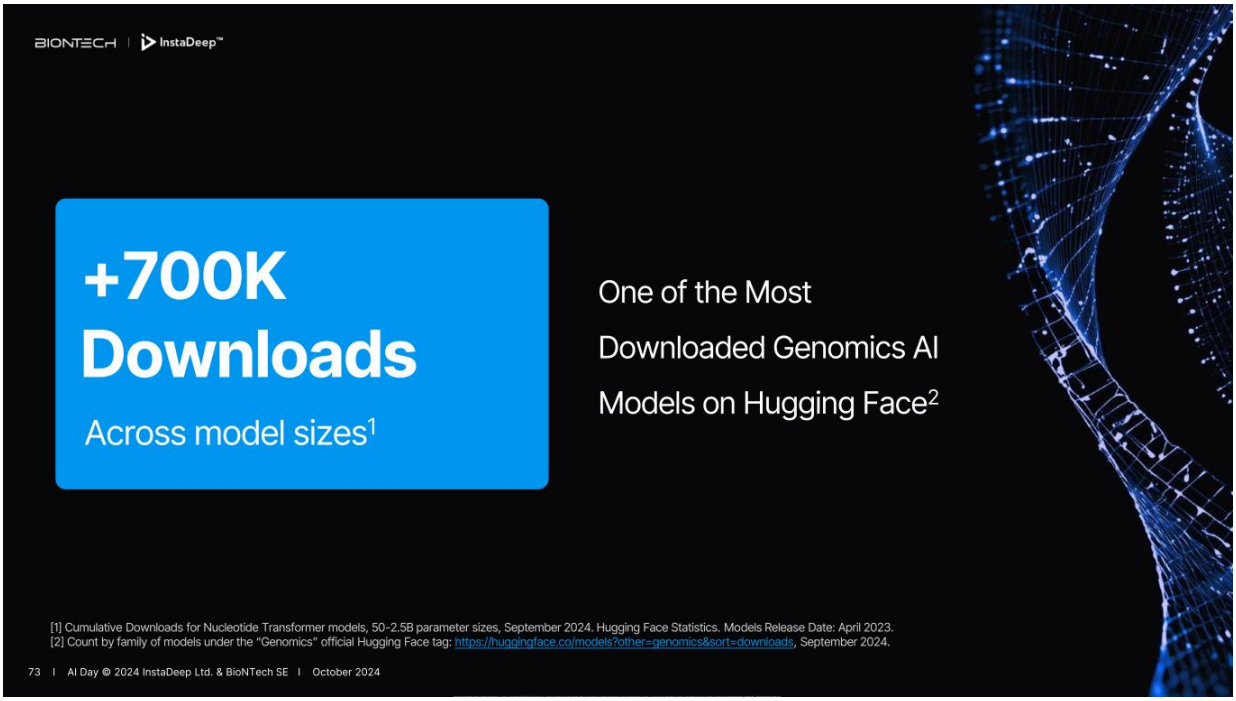
BIONTECH | InstaDeep™

Our Foundation Models for Genomics are State-of-the-Art

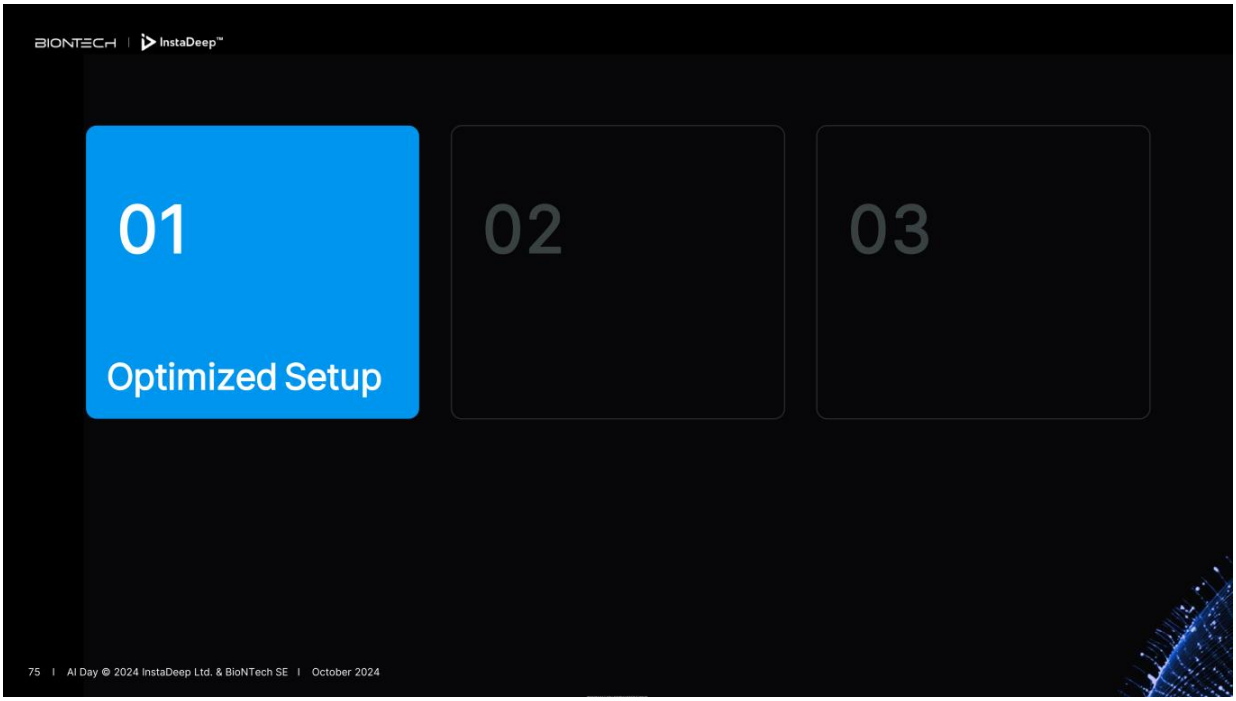
Nucleotide transformer models are state-of-the-art in the space[1]

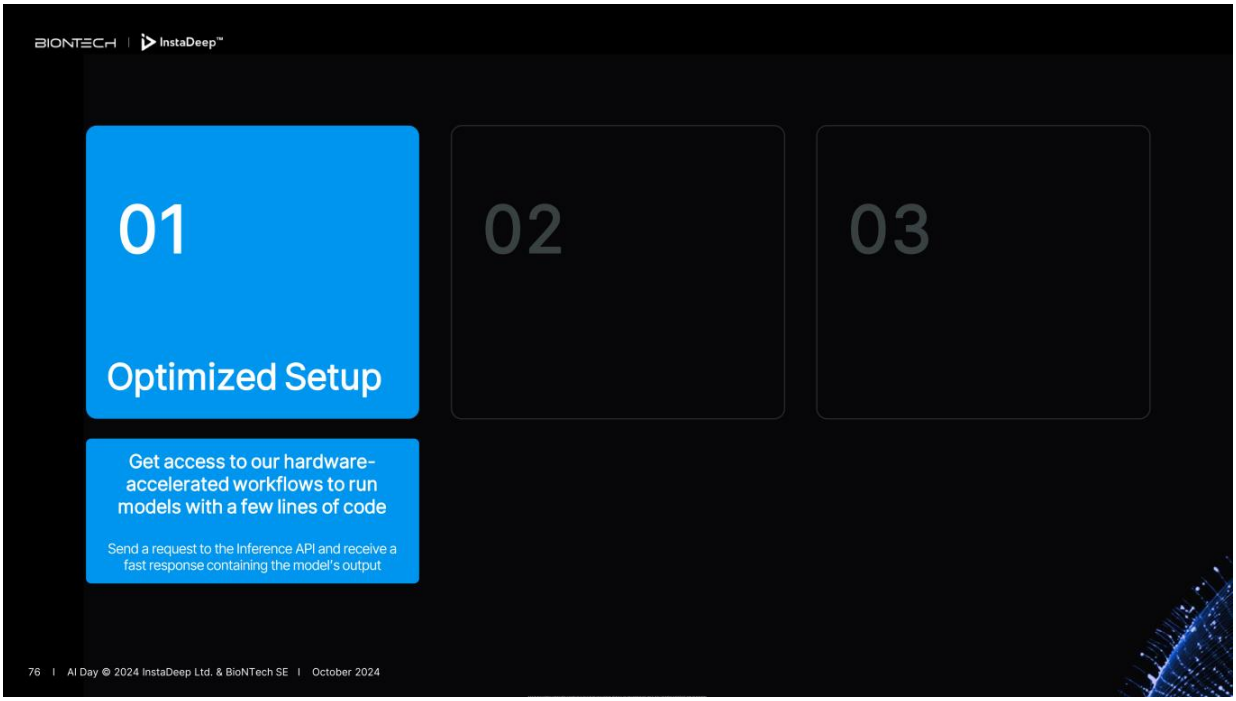


[1] The Nucleotide Transformer: Building and Evaluating Robust Foundation Models for Human Genomics, Dalla-Torre et al., Nat. methods (press)


BIONTECH | InstaDeep™
+700K
Downloads
Across model sizes[1]
One of the Most
Downloaded Genomics AI
Models on Hugging Face[2]
[1] Cumulative Downloads for Nucleotide Transformer models, 50-2.5B parameter sizes, September 2024. Hugging Face Statistics. Models Release Date: April 2023.
[2] Count by family of models under the "Genomics" official Hugging Face tag: https://huggingface.co/models?other=genomics&sort=downloads, September 2024.
73 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

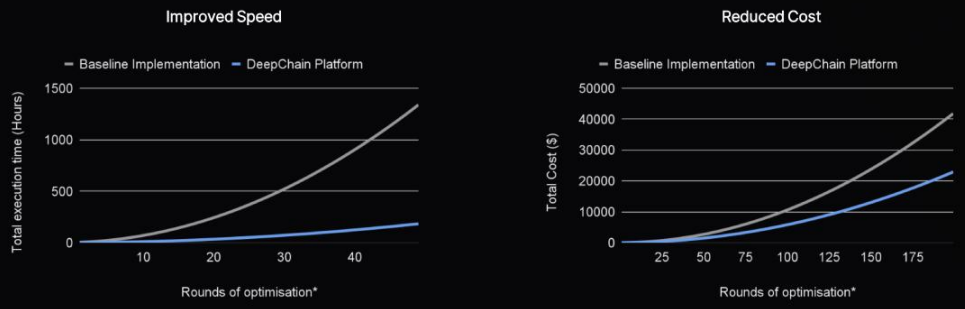

BIONTECH | InstaDeep™
We are releasing capabilities to build and scale with our AI models
74 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
01
Optimized Setup
02
03
75 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

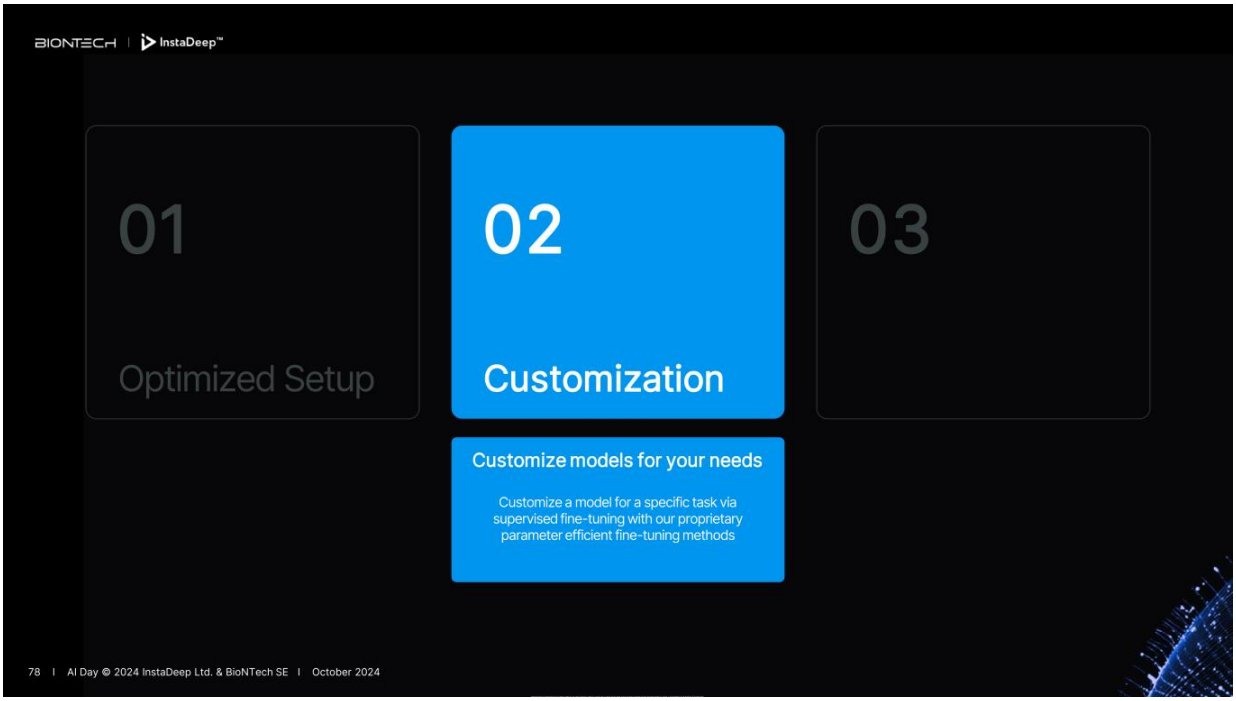

BIONTECH | InstaDeep™
01
Optimized Setup
02
03
Get access to our hardware-accelerated workflows to run models with a few lines of code
Send a request to the Inference API and receive a fast response containing the model's output
76 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

BIONTECH | InstaDeep™

Running inference with DeepChain is 7X faster and 2X cheaper for *in silico* design of regulatory sequences*

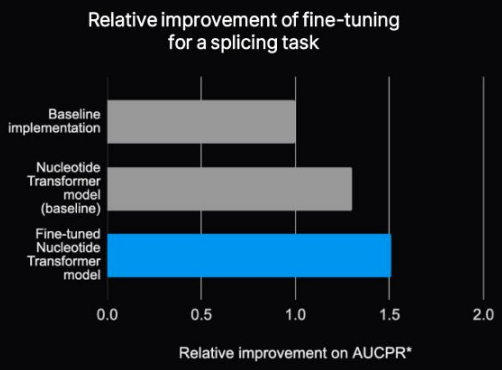


* reference methodology: Jores, T., Tonnies, J., Wrightsman, T. et al. Synthetic promoter designs enabled by a comprehensive analysis of plant core promoters. Nat. Plants 7, 842–855 (2021).
* Test implementation: Sequence length: 6kbp and 2.1kbp sequences, Parameters: --num_indels=8000, --prop_indels=0.5, --random_indels=True, --min_indels_size=2, --max_indel_size=5, --tissue_optimize_idx=1, --opt_metric=increase, --num_rounds=30
* Baseline implementation set-up: 1 NVIDIA V100 Tensor Core GPUs, using published Pytorch implementation available on Hugging Face

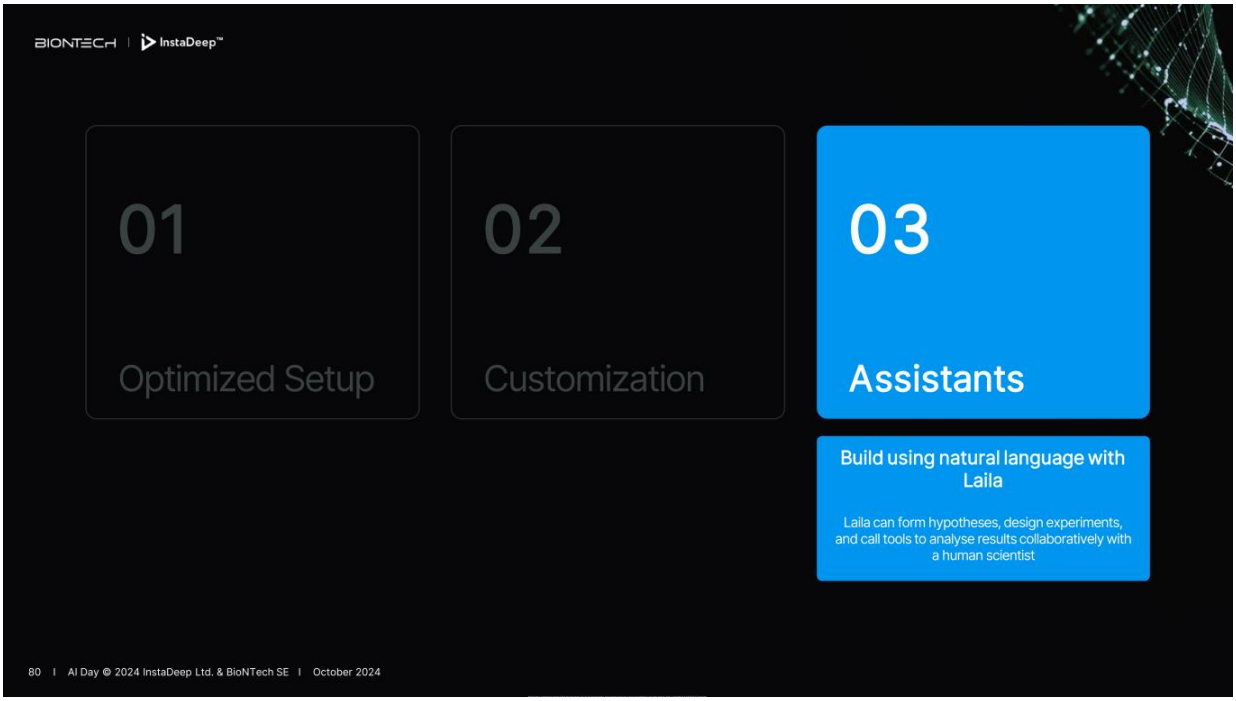

BIONTECH | InstaDeep™
01
Optimized Setup
02
Customization
03
Customize models for your needs
Customize a model for a specific task via supervised fine-tuning with our proprietary parameter efficient fine-tuning methods
78 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

BIONTECH | InstaDeep™

Fine-tuning a model on a specialised data set increased performance ≈1.5X for a splicing prediction use-case



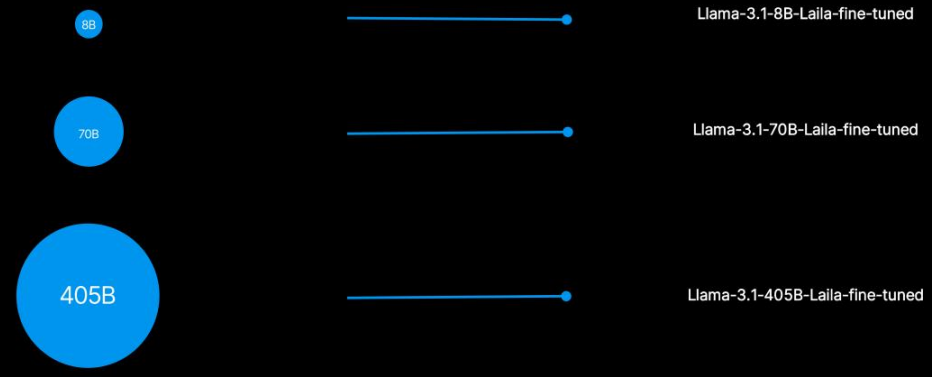
* AUCPR: metric that measures the overall performance of a binary classification model by plotting precision against recall at different threshold settings, providing a more accurate assessment of performance for imbalanced classes.
* Dataset used for customer-specific calculation: Shiraishi Y, Kataoka K, Chiba K, Okada A, Kogure Y, Tanaka H, Ogawa S, Miyano S. A comprehensive characterization of cis-acting splicing-associated variants in human cancer. Genome Res. 2018 Aug;28(8):1111-1125. doi: 10.1101/gr.231951.117. Epub 2018 Jul 16. PMID: 30012835; PMCID: PMC6071634.
* Baseline implementation: Predicting Splicing from Primary Sequence with Deep Learning. Jaganathan, Kishore et al. Cell, Volume 176, Issue 3, 535 - 548.e24


BIONTECH | InstaDeep™
01
Optimized Setup
02
Customization
03
Assistants
Build using natural language with Laila
Laila can form hypotheses, design experiments, and call tools to analyse results collaboratively with a human scientist
80 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
Demo screen
81 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

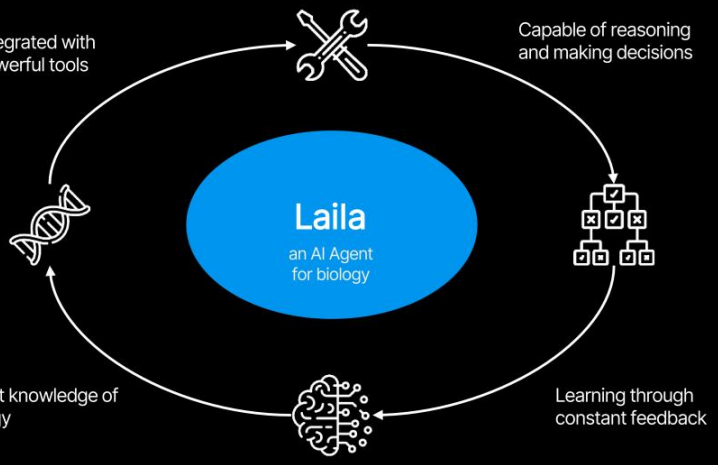
BIONTECH | InstaDeep™

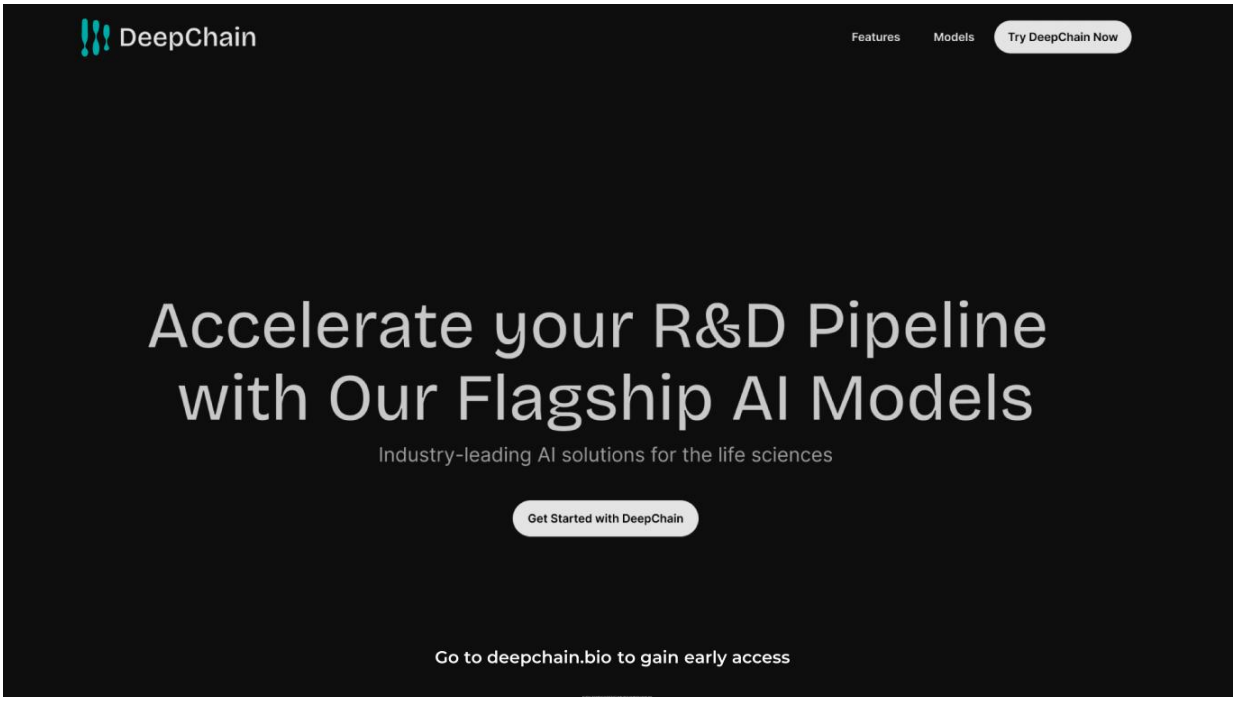
The Laila Series of AI Agents is built with Meta Llama 3.1



BIONTECH | InstaDeep™

Laila is more than just a chatbot...

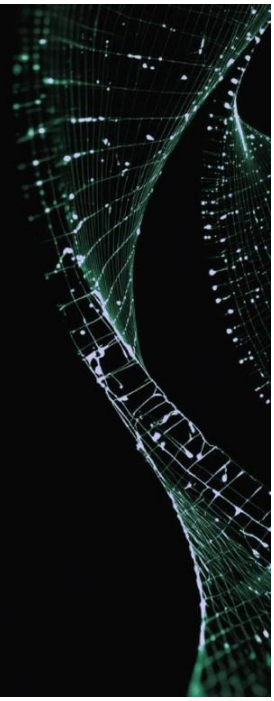



DeepChain
Features
Models
Try DeepChain Now
Accelerate your R&D Pipeline with Our Flagship AI Models
Industry-leading AI solutions for the life sciences
Get Started with DeepChain
Go to deepchain.bio to gain early access

BIONTECH | InstaDeep™

Part II.

Deploying AI Across The Pipeline



BIONTECH | InstaDeep™

Our Goal: Deploying AI end-to-end in our immunotherapy pipeline

AI-first Immunotherapy Platform

1. Immunohistochemistry → 2. DNA/RNA sequencing → 3. Proteomics → 4. Protein Design → 5. Lab Functional Validation

Histology

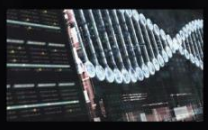
AI computer vision to improve speed and accuracy of tissue labeling

AI Vision

DNA

Sequence analysis and personalized genomic annotations

DNA LLMs



Proteomics

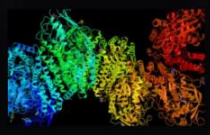
Leveraging AI for target discovery and analysis of the immunological landscape

Protein LLMs

Protein Design

Developing assets for immunotherapy modalities (antibodies, cytokines, TCRs)

Protein LLMs

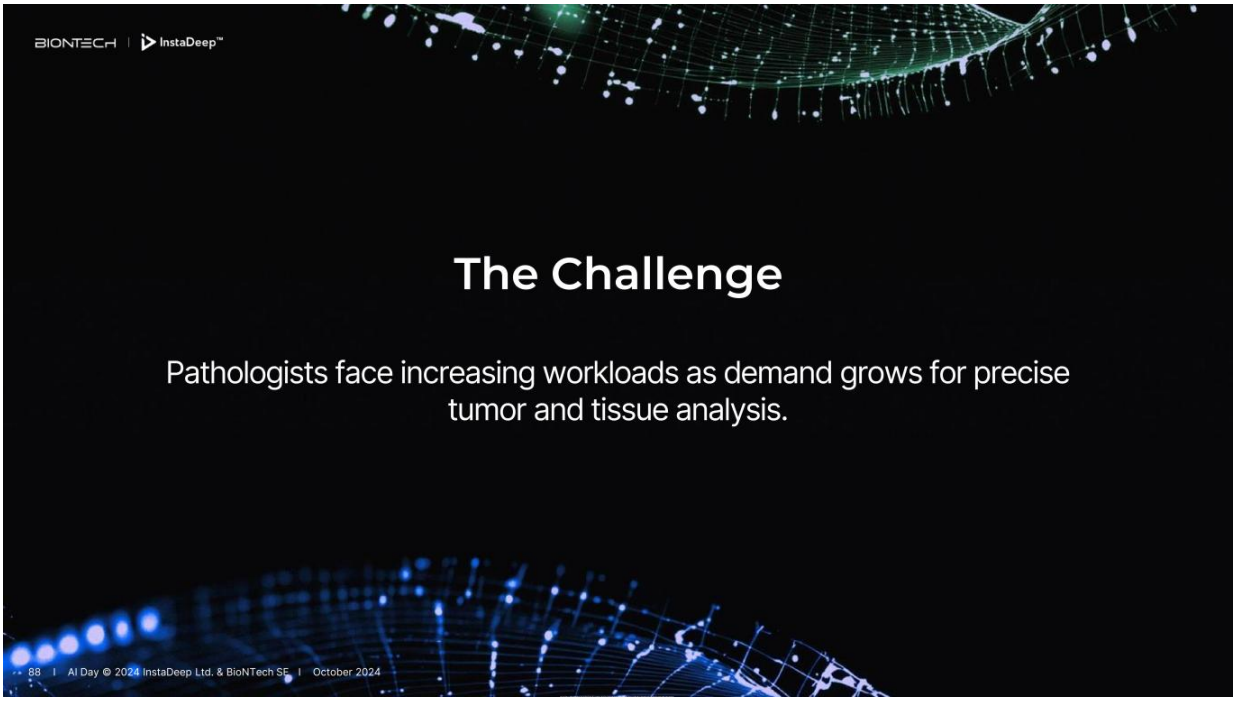


Lab

Instrument automation and quality control

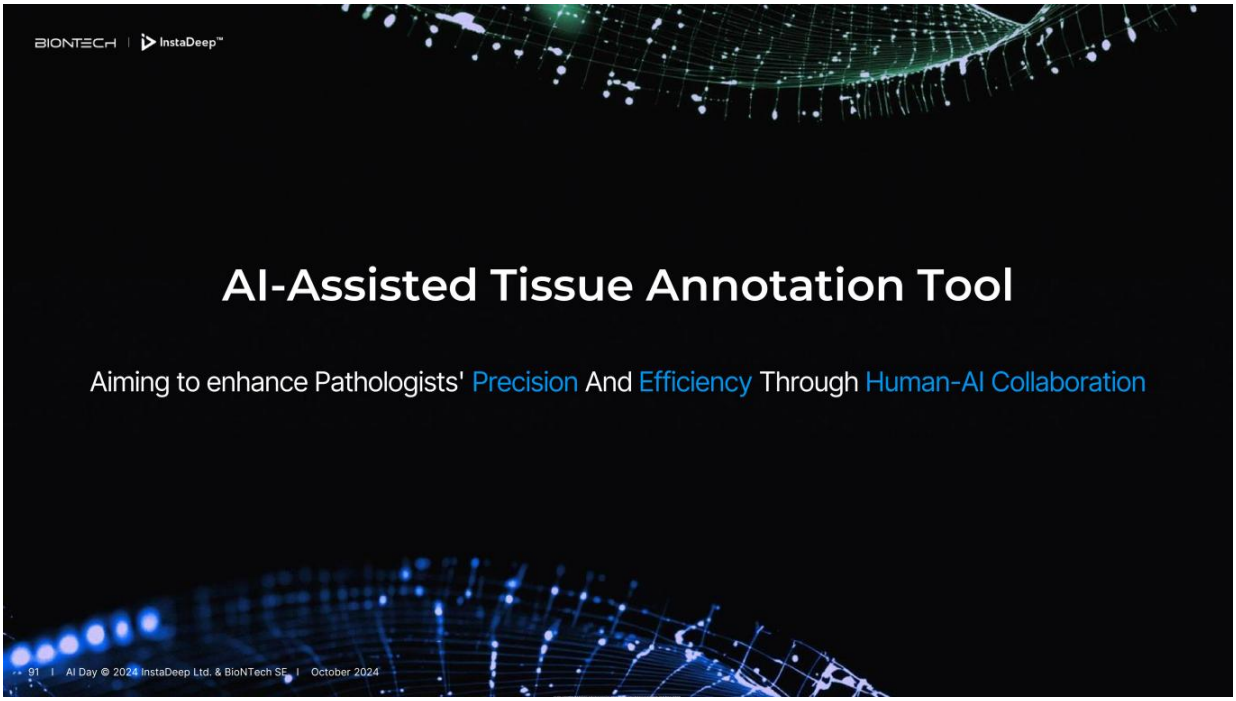
AI Agent

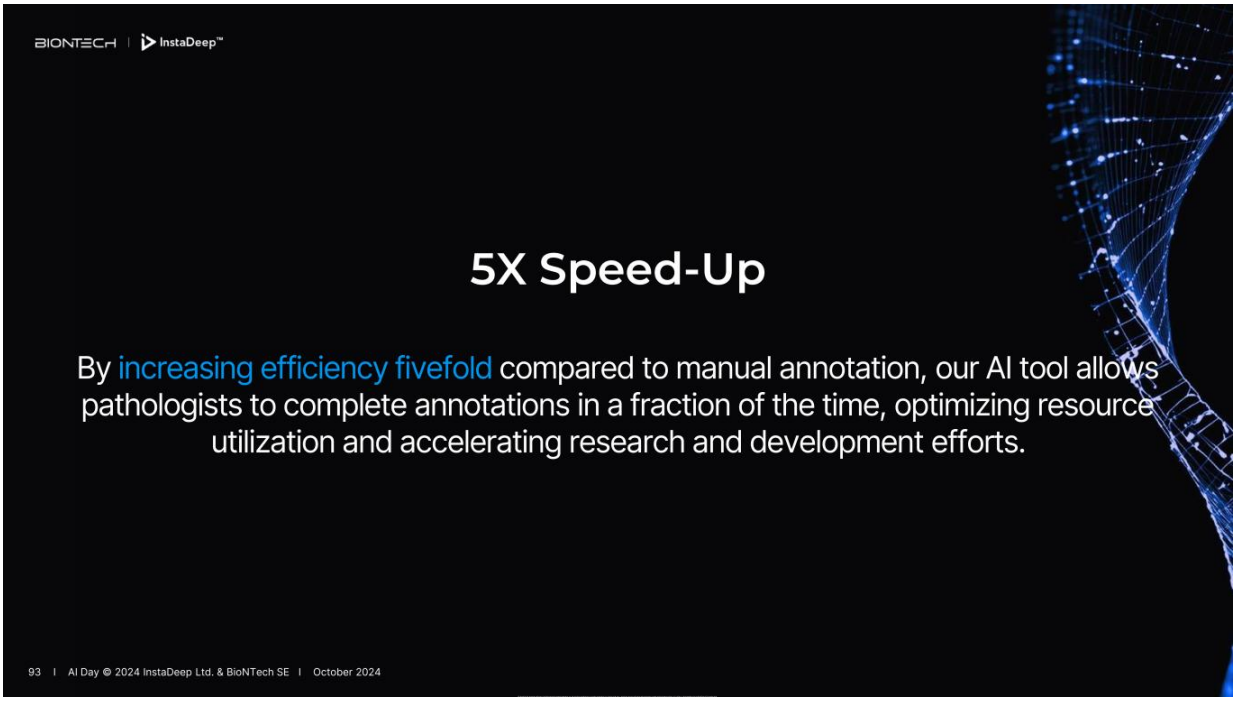

BIONTECH | InstaDeep™
Step 1:
Histology
87 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

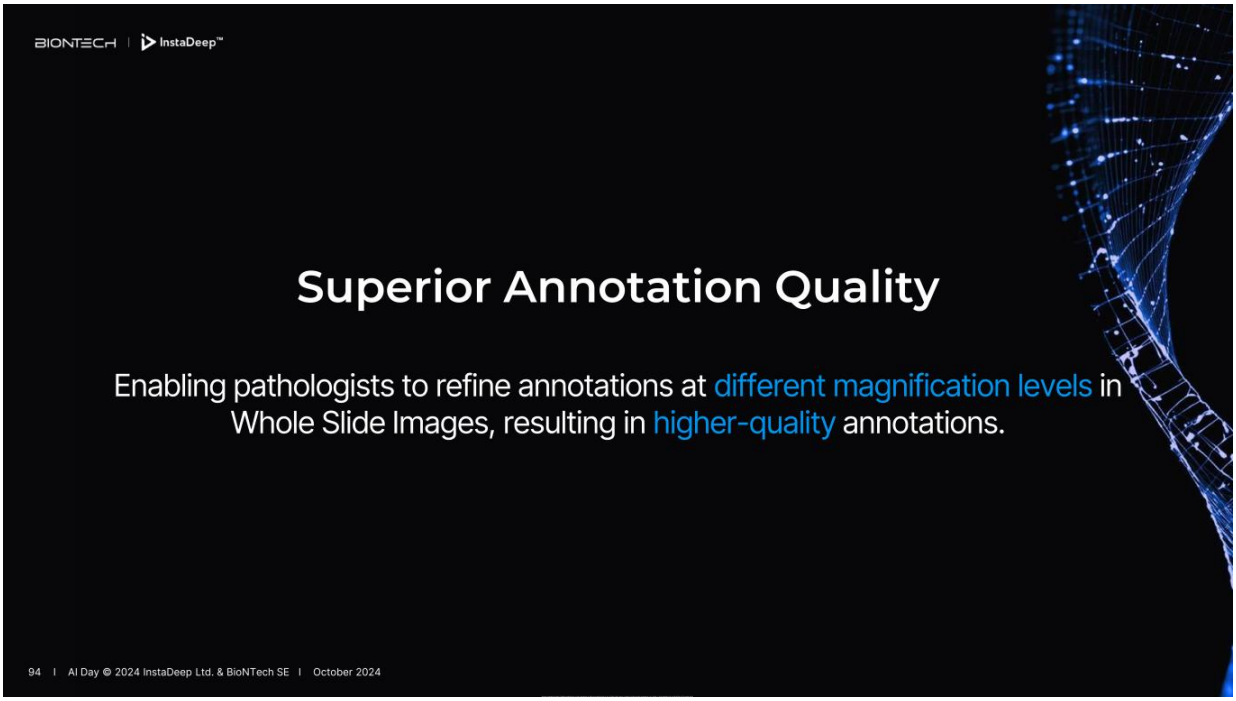

BIONTECH | InstaDeep™
The Challenge
Pathologists face increasing workloads as demand grows for precise tumor and tissue analysis.
88 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


Source: Own Data
2 mm
16451.54, 6953.85 µm
157, 79, 178


BIONTECH | InstaDeep™
Our Approach
Develop tools using state-of-the-art AI to enhance pathologists' workflows
by producing faster and higher-quality annotations.
90 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
AI-Assisted Tissue Annotation Tool
Aiming to enhance Pathologists' Precision And Efficiency Through Human-AI Collaboration
91 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
5X Speed-Up
By increasing efficiency fivefold compared to manual annotation, our AI tool allows pathologists to complete annotations in a fraction of the time, optimizing resource utilization and accelerating research and development efforts.
93 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
Superior Annotation Quality
Enabling pathologists to refine annotations at different magnification levels in Whole Slide Images, resulting in higher-quality annotations.
94 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
Whole Slide Image Segmentation Tool
95 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

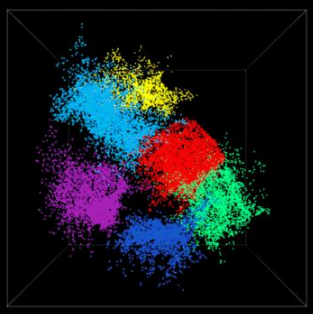
BIONTECH | InstaDeep™

From Classification To Segmentation

Use a state-of-the-art vision foundation model that we train specifically on pathology images.

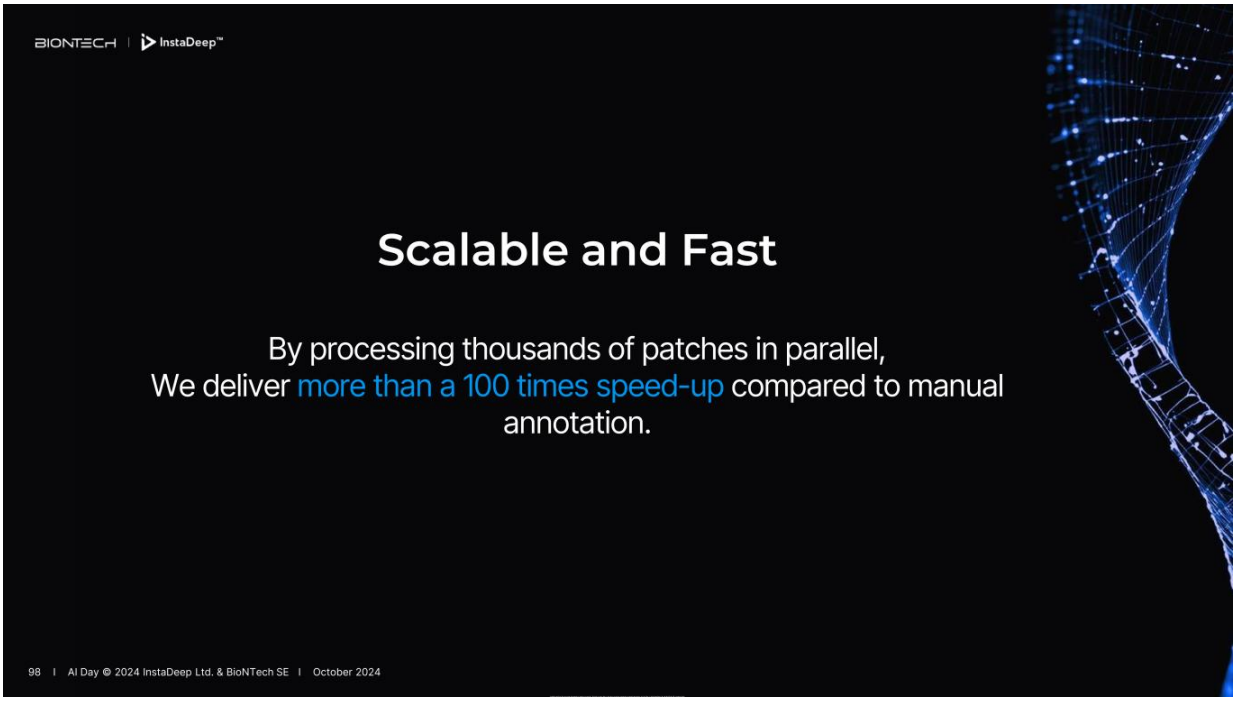
Decompose the image into patches.

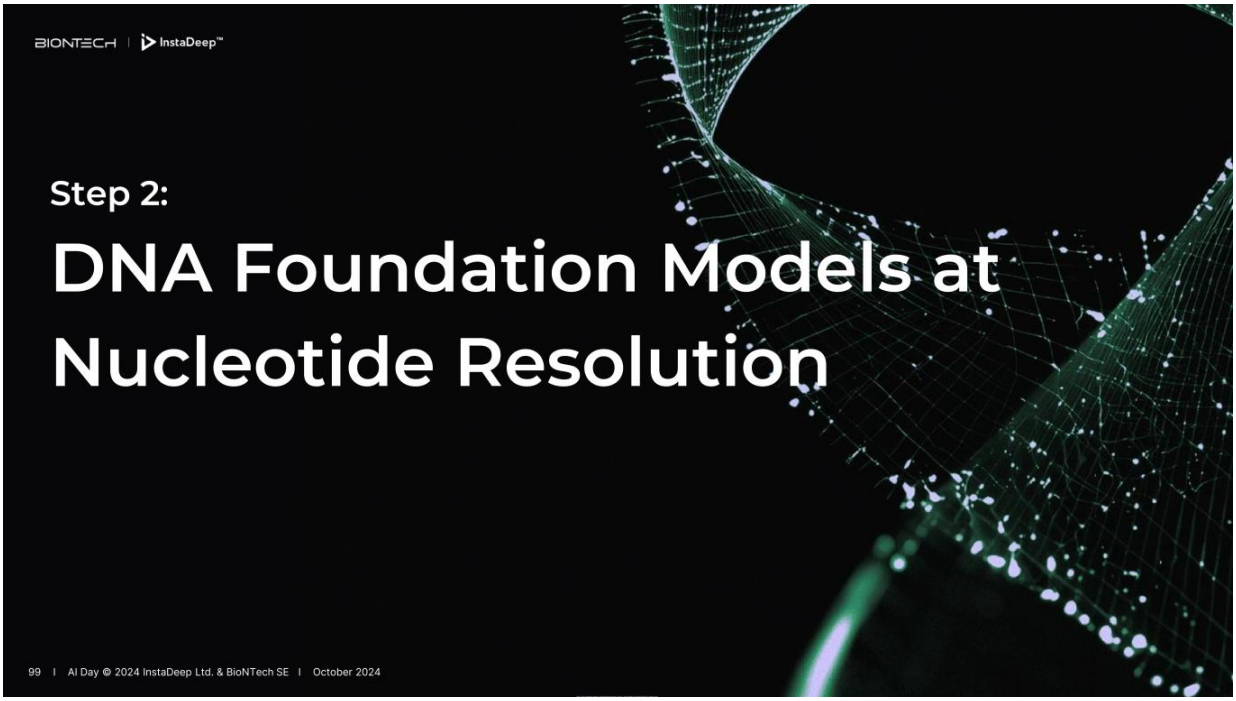
Transform the process from image segmentation to patch classification.



Source: Own Data

Source: Own Data


BIONTECH | InstaDeep™
Scalable and Fast
By processing thousands of patches in parallel,
We deliver more than a 100 times speed-up compared to manual annotation.
98 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
Step 2:
DNA Foundation Models at Nucleotide Resolution
99 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

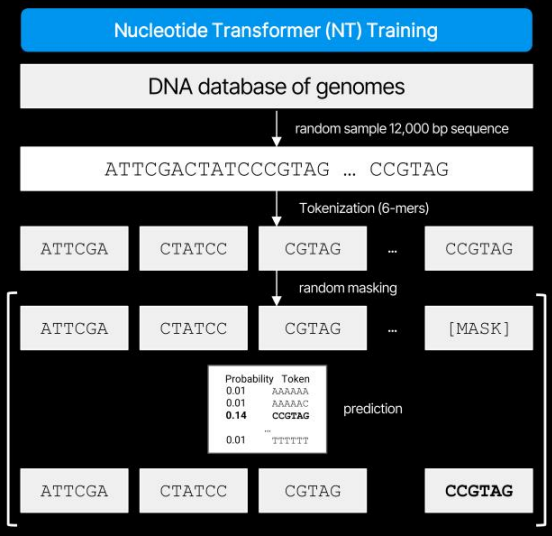
BIONTECH | InstaDeep™

Nucleotide Transformer: Self-Supervised Learning on Genomes

We believe **automated analysis and predictions from genome** sequences have the potential to **transform tomorrow's health care** and **agriculture.**

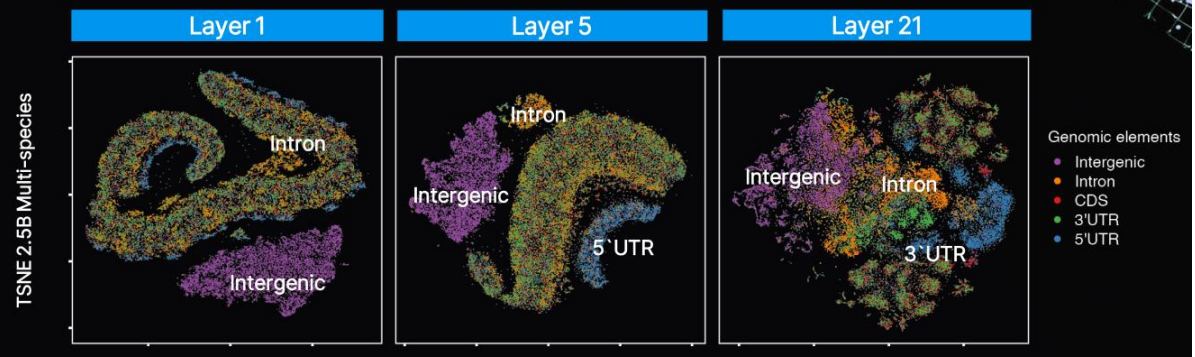
InstaDeep's Nucleotide Transformer Models

- *Architecture*: Masked Language Models (Bert-style training).
- *Datasets*: Trained on 5 datasets from different sizes with inter and intra-species variability from the whole tree of life.
- ***Nucleotide Transformers (NT)***
  - V1: 500M, 1B, 2.5B parameters (2022)
  - V2: 50M, 100M, 250M parameters (2023)
- *Hardware*
  - Cambridge-1 Datacenter (collaboration with Nvidia)
  - TPUv4-1024 Pod (collaboration with Google Cloud)



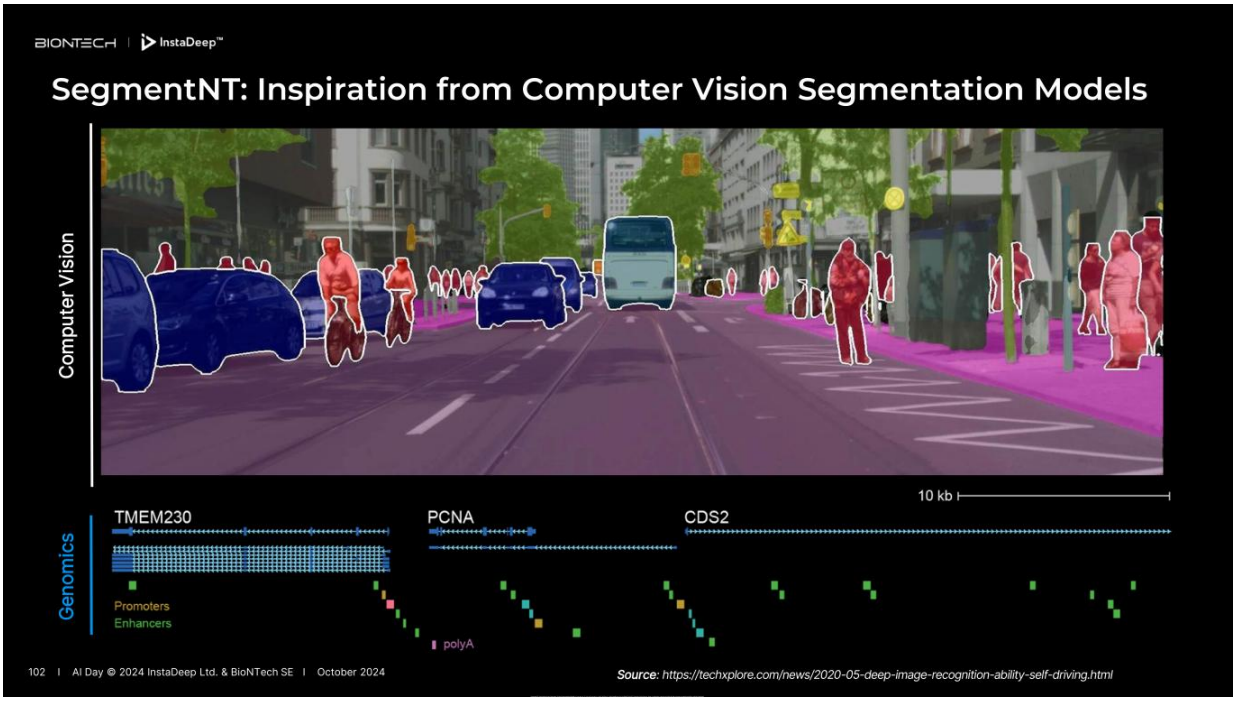
BIONTECH | InstaDeep™

NT Acquires Genomics Knowledge During Pre-Training



Even without supervision, information about **genomic sequence features is learned** in the "sequence representation"

Source: Dalla-Torre et al., Nat Methods 2024 (in press)


BIONTECH | InstaDeep™
SegmentNT: Inspiration from Computer Vision Segmentation Models
Computer Vision
10 kb
Genomics
TMEM230
PCNA
CDS2
Promoters
Enhancers
polyA
102 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024
Source: https://techxplore.com/news/2020-05-deep-image-recognition-ability-self-driving.html

BIONTECH | InstaDeep™

SegmentNT: Annotating the Genome at Nucleotide Resolution

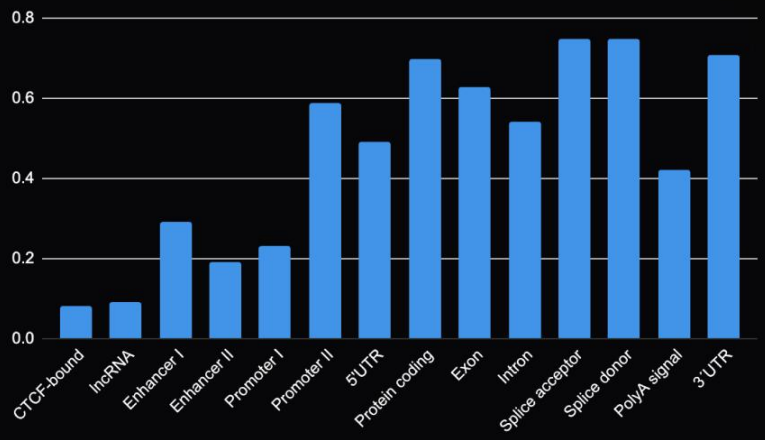
We fine-tuned the Nucleotide Transformers on **2.5M** high-quality **gene and regulatory elements annotations**.

14 annotations per nucleotide (e.g. 700,000 predictions at 50kbp)

BIONTECH | InstaDeep™

SegmentNT: Annotating the Genome at Nucleotide Resolution

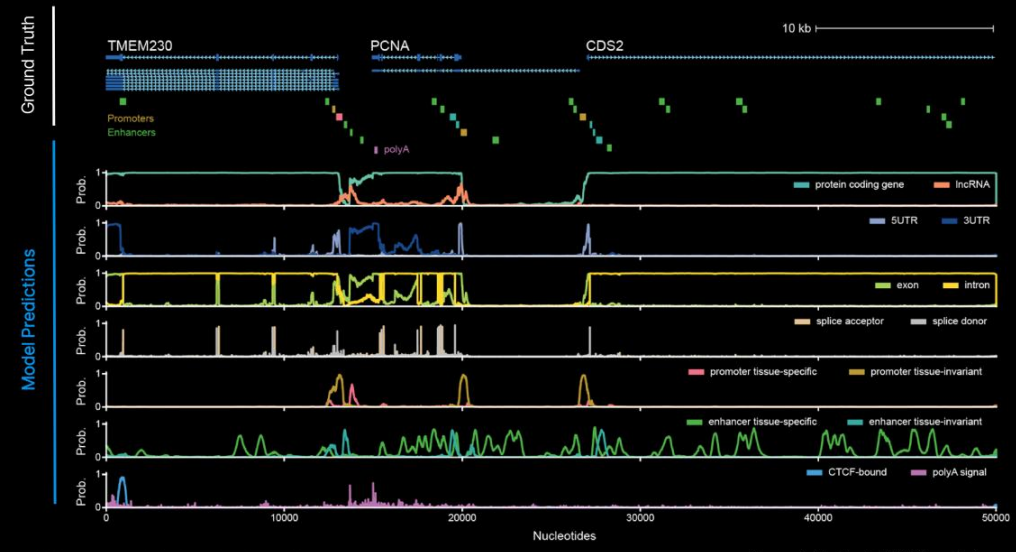
The resulting SegmentNT model segments sequences up to **50kbp** with **state-of-the-art performance** for **splicing**, **gene finding** and **regulatory element detection.**



Source: adapted from de Almeida et al., 2024 (in revision)

BIONTECH | InstaDeep™

SegmentNT: 700,000 Accurate Predictions over 50kbp in less than a second

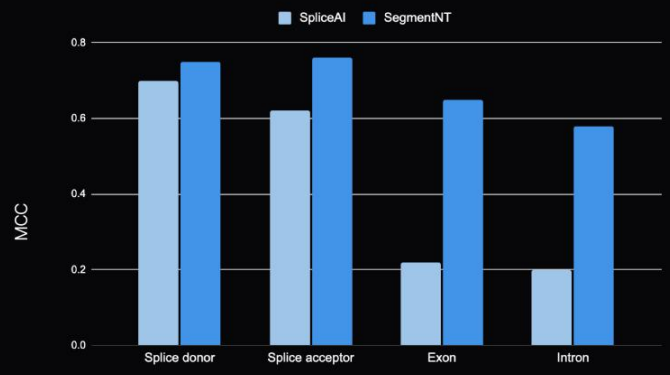


Source: de Almeida et al., 2024 (in revision)

BIONTECH | InstaDeep™

SegmentNT is State-of-the-Art for Canonical Splicing Detection

Splicing is a biological process that **removes non-coding sequences** (introns) from a primary messenger RNA (mRNA) transcript and joins the coding sequences (exons) together to create a mature mRNA. **Dysregulated splicing** can be a **vulnerability in cancers**. SegmentNT **outperforms state-of-the-art SpliceAI** for splicing event detection.



Test performance for splicing detection over full chromosomes on human reference genome

Source: de Almeida et al., 2024 (in revision)

BIONTECH | InstaDeep™

Alternative Splicing Events Detection with SegmentNT

Alternative splicing events can disrupt protein production and cancer pathways and is associated with cancer development. We finetuned segmentNT to identify tumor antigen candidates from alternative splicing events, which represent potential targets for personalized cancer immunotherapies. After finetuning, SegmentNT can accurately predict alternative splicing events in cancer data.

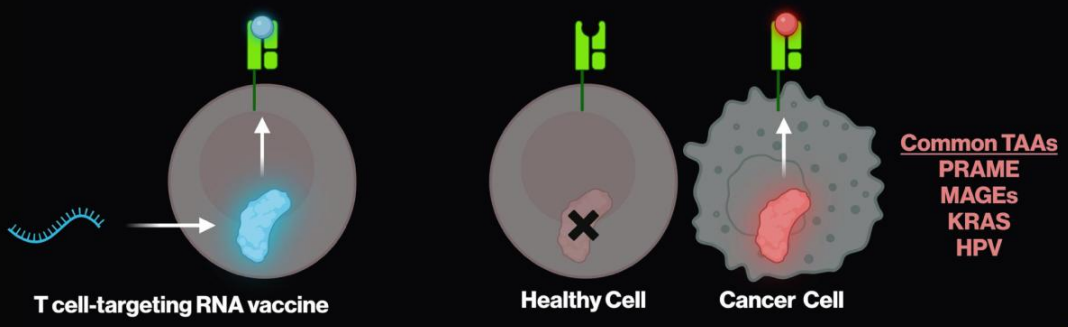
Test performance for 2,000 alternative splicing detection over cancer data (TCGA LUAD data)

Source: https://www.cancer.gov/tcga (data)

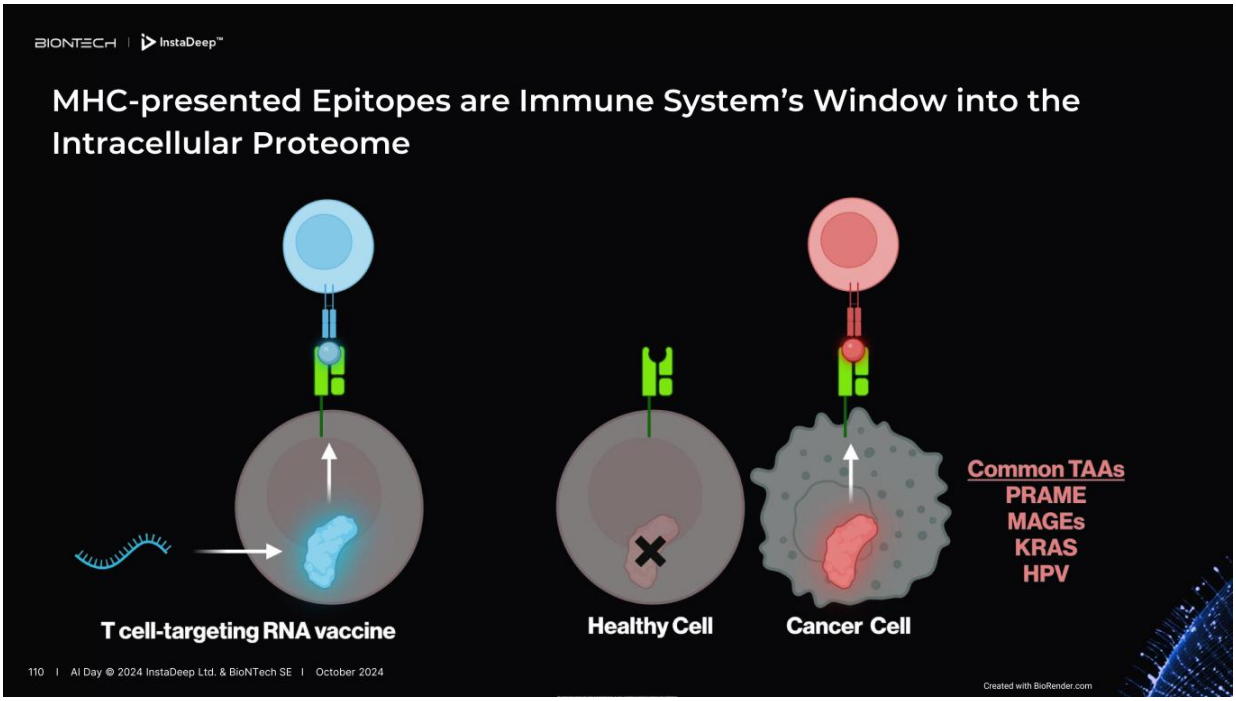

BIONTECH | InstaDeep™
Step 3:
AI-Enhanced Proteomics for Target Discovery
108 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

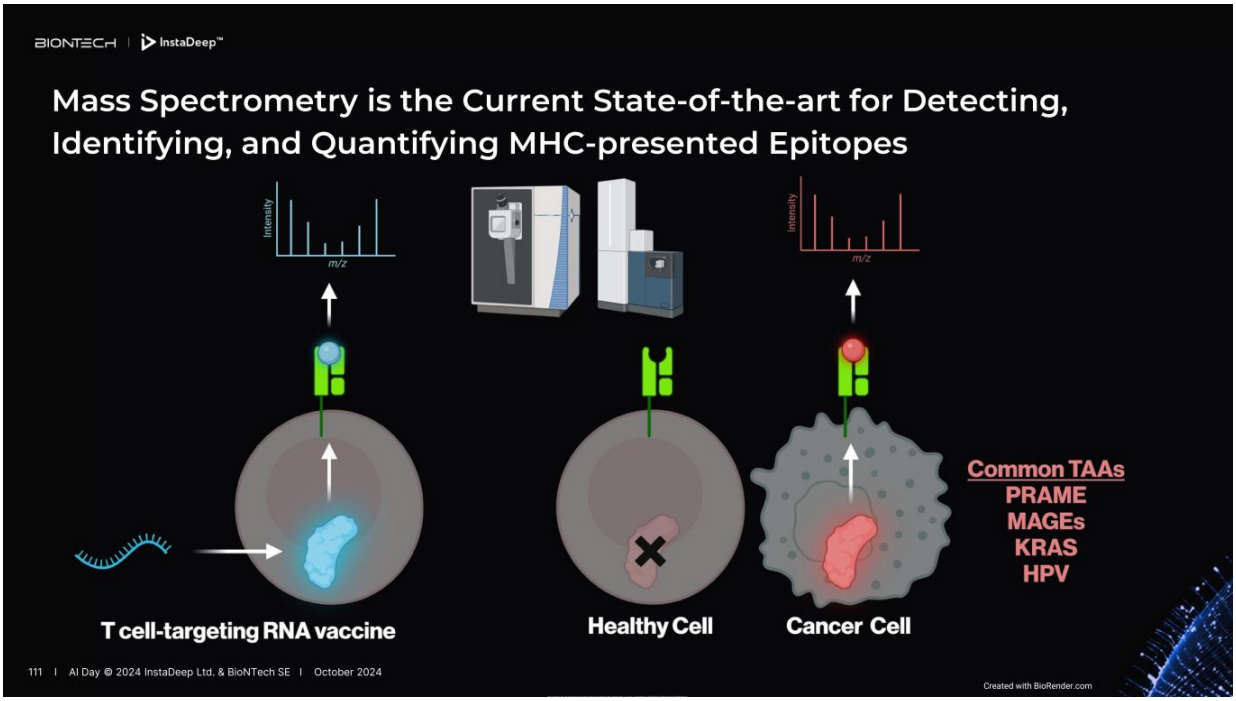
BIONTECH | InstaDeep™

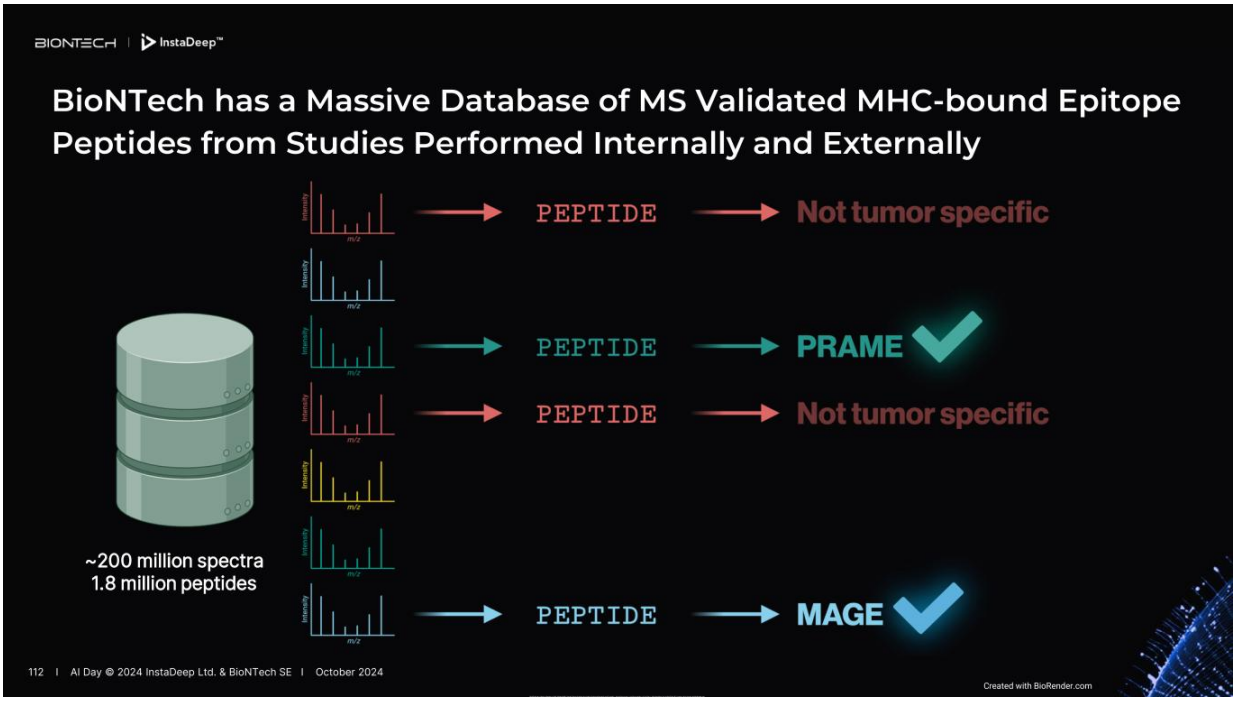
Intracellular Proteins are Processed and Presented on MHC Complexes



Created with BioRender.com


BIONTECH | InstaDeep™
MHC-presented Epitopes are Immune System's Window into the Intracellular Proteome
T cell-targeting RNA vaccine
Healthy Cell
Cancer Cell
Common TAAs
PRAME
MAGEs
KRAS
HPV
110 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024
Created with BioRender.com

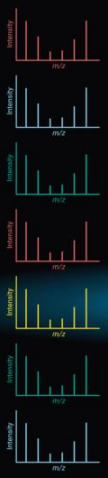

BIONTECH | InstaDeep™
Mass Spectrometry is the Current State-of-the-art for Detecting, Identifying, and Quantifying MHC-presented Epitopes
Intensity
m/z
Intensity
m/z
Common TAAs
PRAME
MAGEs
KRAS
HPV
T cell-targeting RNA vaccine
Healthy Cell
Cancer Cell
111 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024
Created with BioRender.com


BIONTECH | InstaDeep™
BioNTech has a Massive Database of MS Validated MHC-bound Epitope Peptides from Studies Performed Internally and Externally
PEPTIDE
Not tumor specific
PEPTIDE
PRAME
PEPTIDE
Not tumor specific
PEPTIDE
MAGE
~200 million spectra
1.8 million peptides
112 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024
Created with BioRender.com

BIONTECH | InstaDeep™

With the Power of AI, we can Dig Deeper into our Data to Identify Novel Therapeutic Targets

~200 million spectra
>>1.8 million peptides



PEPTIDE → Not tumor specific

PEPTIDE → PRAME ✓

PEPTIDE → Not tumor specific

PEPTIDE → Novel Target ✓

PEPTIDE → MAGE ✓

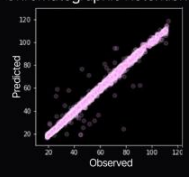
Created with BioRender.com

BIONTECH | InstaDeep™

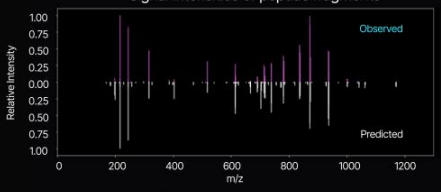
AI Maximizes our Ability to Discover Novel Cancer Targets

Up to 200% increase in recovered peptide IDs

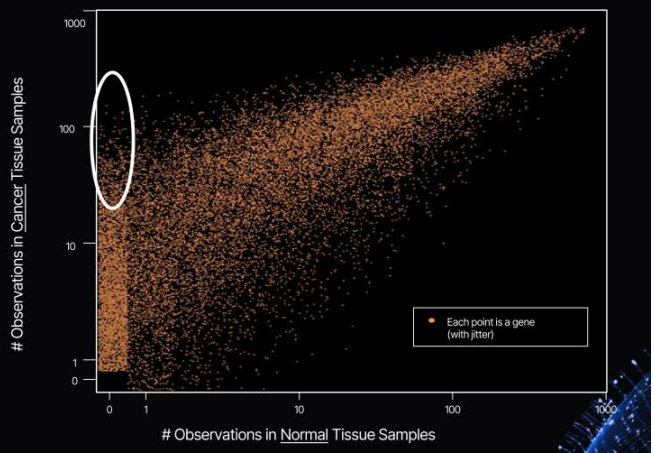
Chromatographic Retention Time



Signal intensities of peptide fragments



Identification of novel, tumor-specific peptides



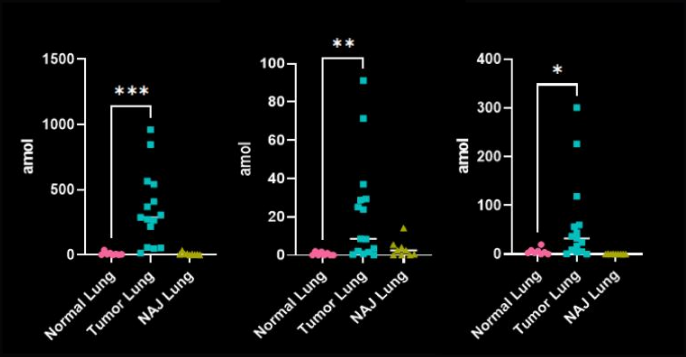
Source : Internal

Source : Internal

BIONTECH | InstaDeep™

Targets are Validated with High-precision Mass Spectrometry

Lung Squamous Cell Carcinoma:
Example peptides validated using synthetic controls



Source : Internal

Validated targets are candidates for TCR-based therapies

Underway: Novel *in silico* approaches to discover and enhance TCRs

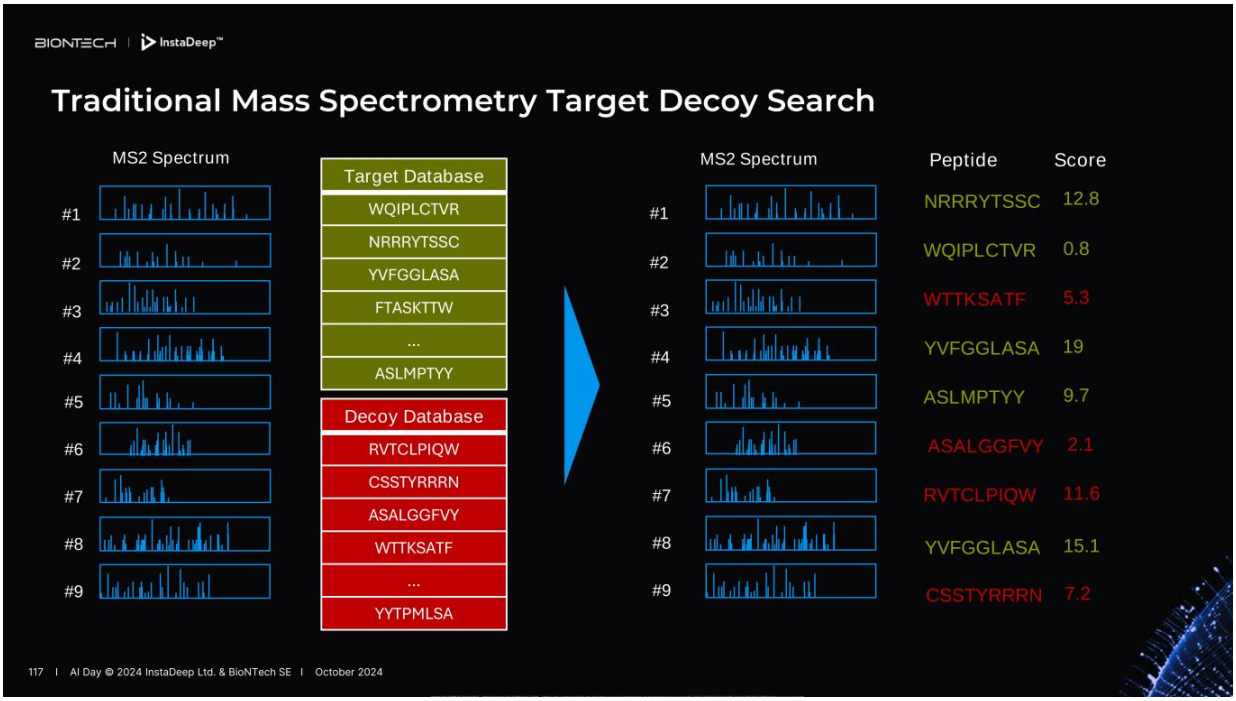
BIONTECH | InstaDeep™

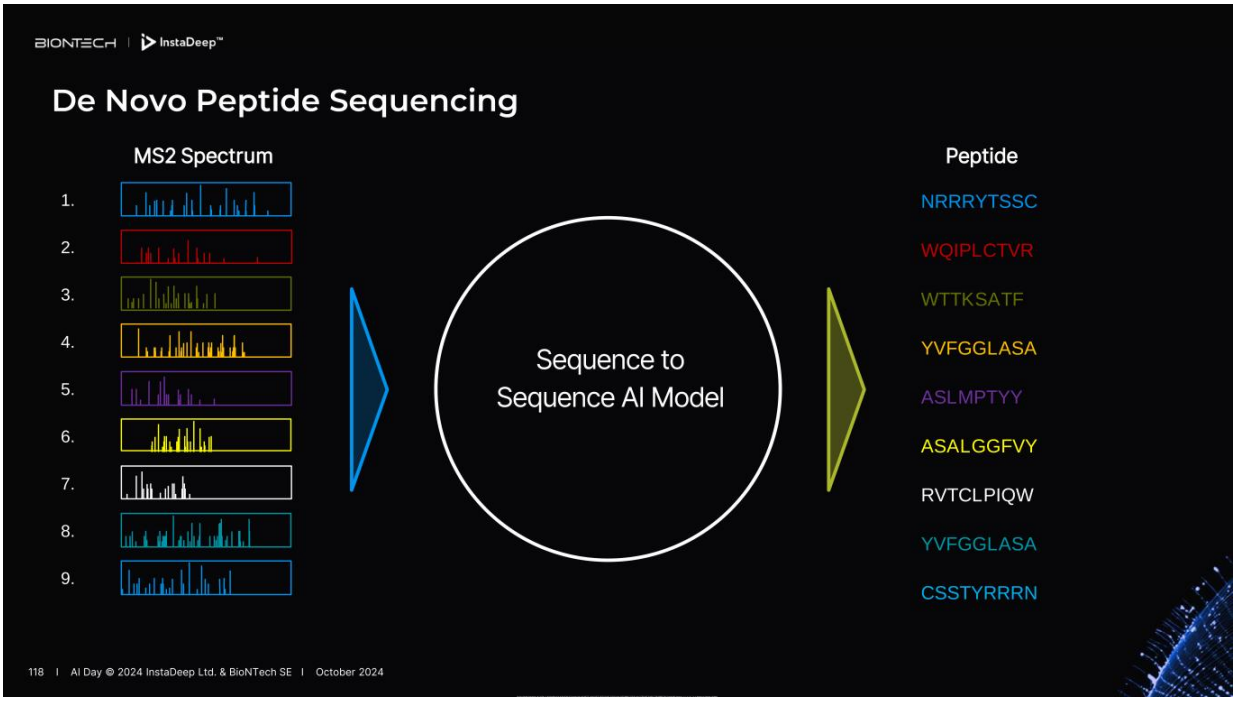
The Challenge of HLA Mass Spectrometry

55-75% of data cannot be mapped to a known human peptide

LncRNAs
Circular RNAs
Non-canonical open reading frames
Endogenous retroviruses
Post-translational modifications
??

Canonical human protein sequences


BIONTECH | InstaDeep™
Traditional Mass Spectrometry Target Decoy Search
MS2 Spectrum
#1
#2
#3
#4
#5
#6
#7
#8
#9
Target Database
WQIPLCTVR
NRRRYTSSC
YVFGGLASA
FTASKTTW
...
ASLMPTYY
Decoy Database
RVTCLPIQW
CSSTYRRRN
ASALGGFVY
WTTKSATF
...
YYTPMLSA
MS2 Spectrum
Peptide
Score
#1 NRRRYTSSC 12.8
#2 WQIPLCTVR 0.8
#3 WTTKSATF 5.3
#4 YVFGGLASA 19
#5 ASLMPTYY 9.7
#6 ASALGGFVY 2.1
#7 RVTCLPIQW 11.6
#8 YVFGGLASA 15.1
#9 CSSTYRRRN 7.2
117 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
De Novo Peptide Sequencing
MS2 Spectrum
1.
2.
3.
4.
5.
6.
7.
8.
9.
Sequence to Sequence AI Model
Peptide
NRRRYTSSC
WQIPLCTVR
WTTKSATF
YVFGGLASA
ASLMPTYY
ASALGGFVY
RVTCLPIQW
YVFGGLASA
CSSTYRRRN
118 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

BIONTECH | InstaDeep™

InstaNovo - *De Novo* Peptide Sequencing with Deep Learning

The approach

De novo peptide sequencing using deep learning. No database needed.

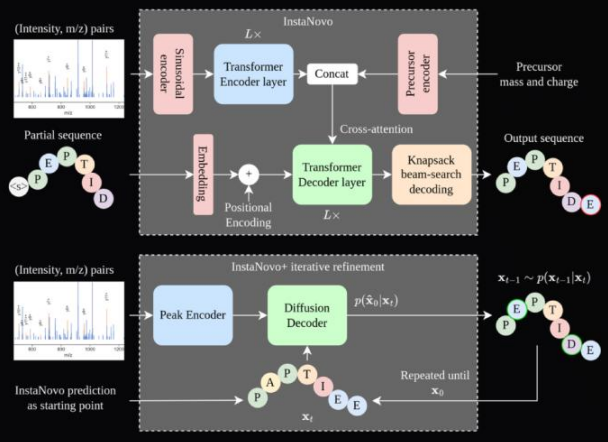
The dataset

Model trained on **28 million labeled spectra** matched to **742 thousand human peptides** from ProteomeTools project.

The models

InstaNovo
Autoregressive encoder-decoder transformer model with special MS2 spectrum encoder

InstaNovo+
Multinomial diffusion model to further improve performance using iterative refinement



Source: De novo peptide sequencing with InstaNovo: Accurate, database-free peptide identification for large scale proteomics experiments (https://www.biorxiv.org/content/10.1101/2023.08.30.555055v3)

BIONTECH | InstaDeep™

InstaNovo - *De Novo* Peptide Sequencing with Deep Learning

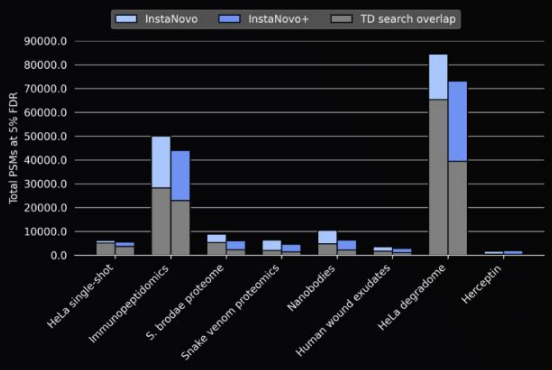
The results

- InstaNovo has performed well across most datasets
- Increases PSM rate in HeLa proteomes
- Expanded an immunopeptidomics dataset by 42%
- Found peptides from individual-specific mutations, splice variants, and post-translational modifications.
- Discovered new HLA peptides in immunopeptidome experiments

Preprint and code available

Preprint is available on **BioRxiv**
https://www.biorxiv.org/content/10.1101/2023.08.30.555055v3

Code is available on **GitHub**
https://github.com/instadeepai/instanovo



Source: De novo peptide sequencing with InstaNovo: Accurate, database-free peptide identification for large scale proteomics experiments (https://www.biorxiv.org/content/10.1101/2023.08.30.555055v3)


BIONTECH | InstaDeep™
Step 4:
Protein Design:
RiboMab™ Platform
121 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

BIONTECH | InstaDeep™

Press Release

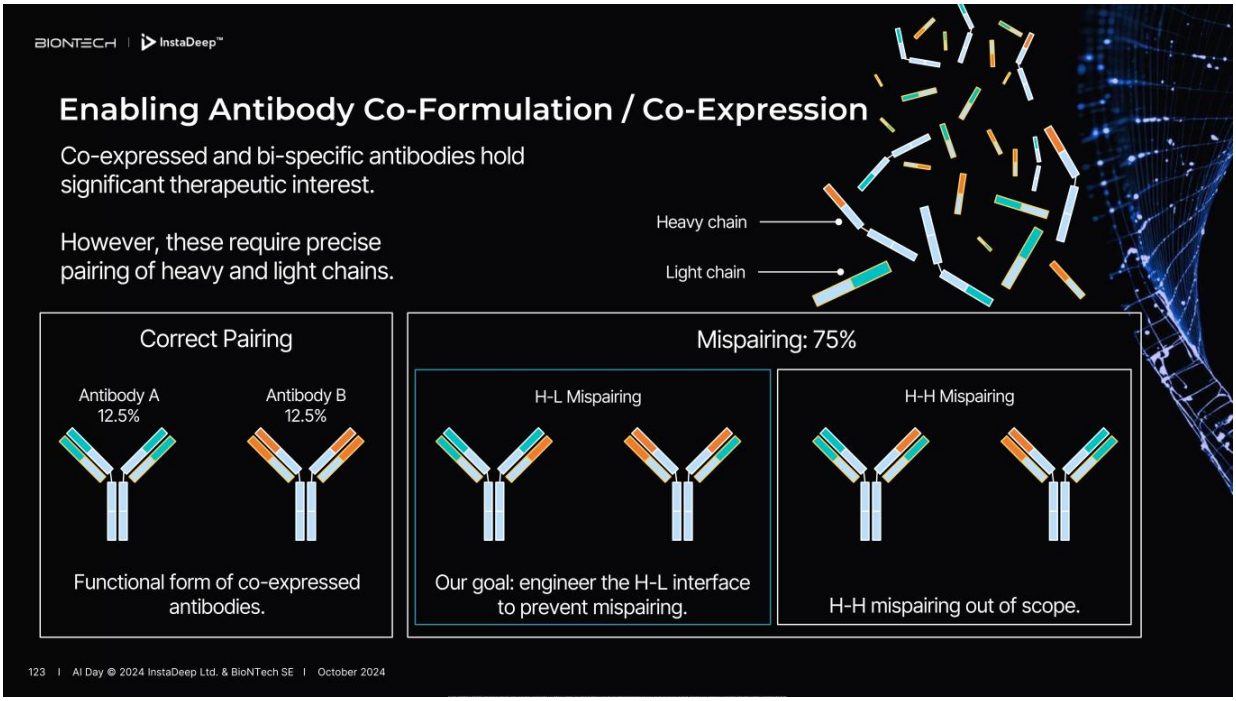
BioNTech and InstaDeep Announce Strategic Collaboration and Form AI Innovation Lab to Develop Novel Immunotherapies

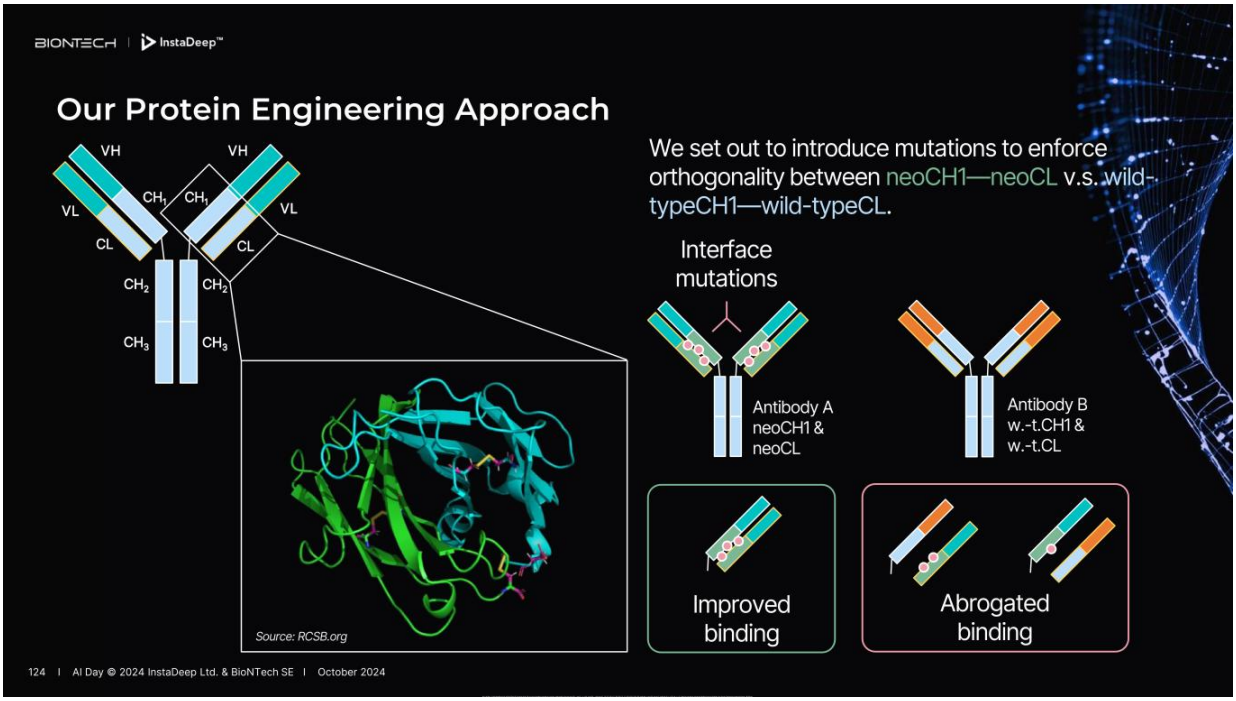
25 November 2020

The strategic collaboration will focus on three core areas:

- **Novel Drug Design:** BioNTech is advancing a pipeline of novel mRNA-based vaccines and therapeutics and will apply InstaDeep's DeepChain™ protein design platform to engineer new mRNA sequences for protein targets, including for its RiboMab™ and RiboCytokine™ platforms, which use messenger RNA to encode antibodies and cytokines *in vivo*.

Source: https://investors.biontech.de/news-releases/news-release-details/biontech-and-instadeep-announce-strategic-collaboration-and-form/


BIONTECH | InstaDeep™
Enabling Antibody Co-Formulation / Co-Expression
Co-expressed and bi-specific antibodies hold significant therapeutic interest.
However, these require precise pairing of heavy and light chains.
Heavy chain
Light chain
Correct Pairing
Antibody A 12.5%
Antibody B 12.5%
Functional form of co-expressed antibodies.
Mispairing: 75%
H-L Mispairing
Our goal: engineer the H-L interface to prevent mispairing.
H-H Mispairing
H-H mispairing out of scope.
123 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
Our Protein Engineering Approach
VH
VH
VL
CH1
CH1
VL
CL
CL
CH2
CH2
CH3
CH3
Source: RCSB.org
We set out to introduce mutations to enforce orthogonality between neoCH1—neoCL v.s. wild-typeCH1—wild-typeCL.
Interface mutations
Antibody A neoCH1 & neoCL
Antibody B w.-t.CH1 & w.-t.CL
Improved binding
Abrogated binding
124 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

BIONTECH | InstaDeep™

Finding Optimal Orthogonal Mutations

A combinatorial multi-objective optimisation problem we could solve thanks to our DeepChain™ platform and an efficient *in silico – in vitro* feedback loop.

Binding energy estimations

For all correctly paired and mispaired complexes.

Mitigation of thermostability changes

For all both heavy and light chains.

Structural modelling

Of each interface mutation.

Key interaction understanding

In depth knowledge of the physics of the interface.

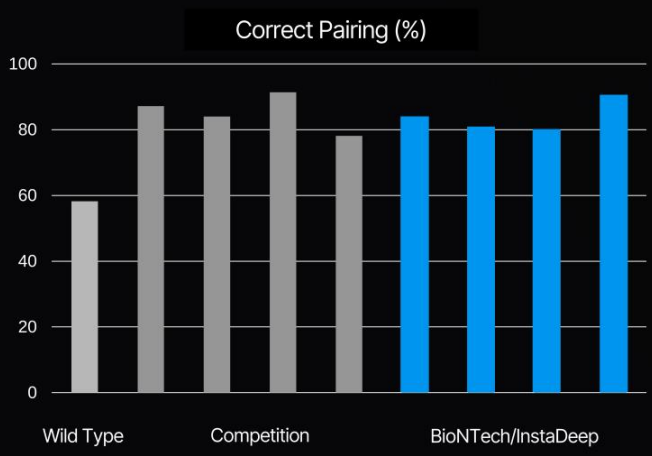
BIONTECH | InstaDeep™

RiboMab™ Results obtained with DeepChain™

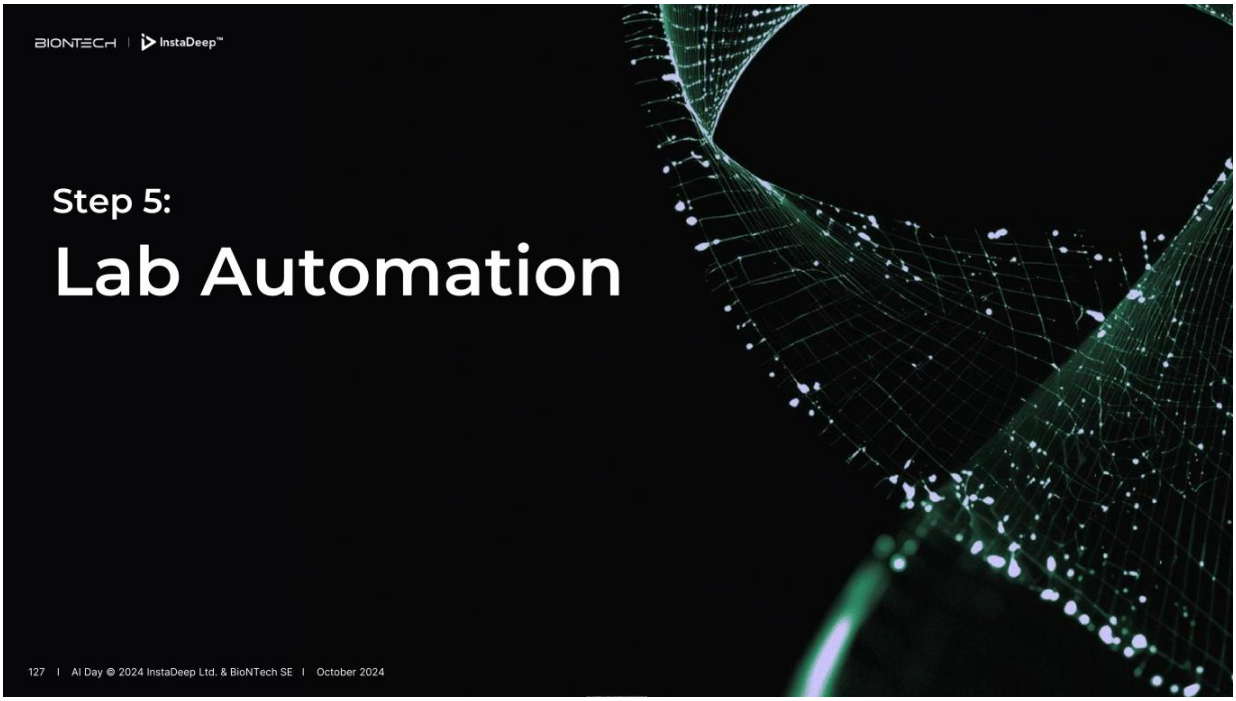
>90% correct pairing

Matching the best patented designs on the market.

Functional activity of antibodies confirmed.

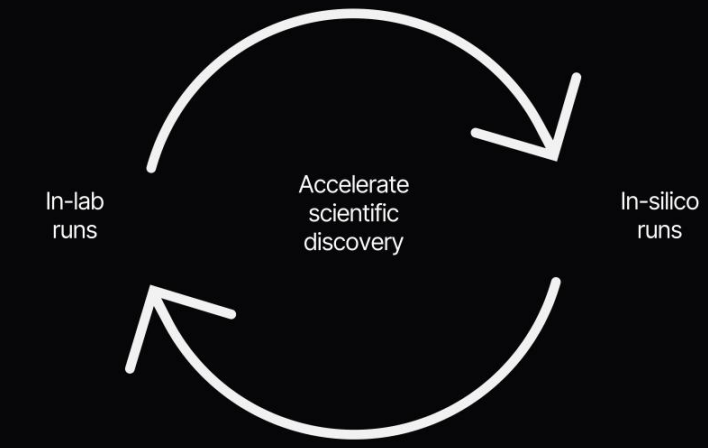


Source: Internal Data.


BIONTECH | InstaDeep™
Step 5:
Lab Automation
127 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

BIONTECH | InstaDeep™

Lab Automation Could Transform Research & Development

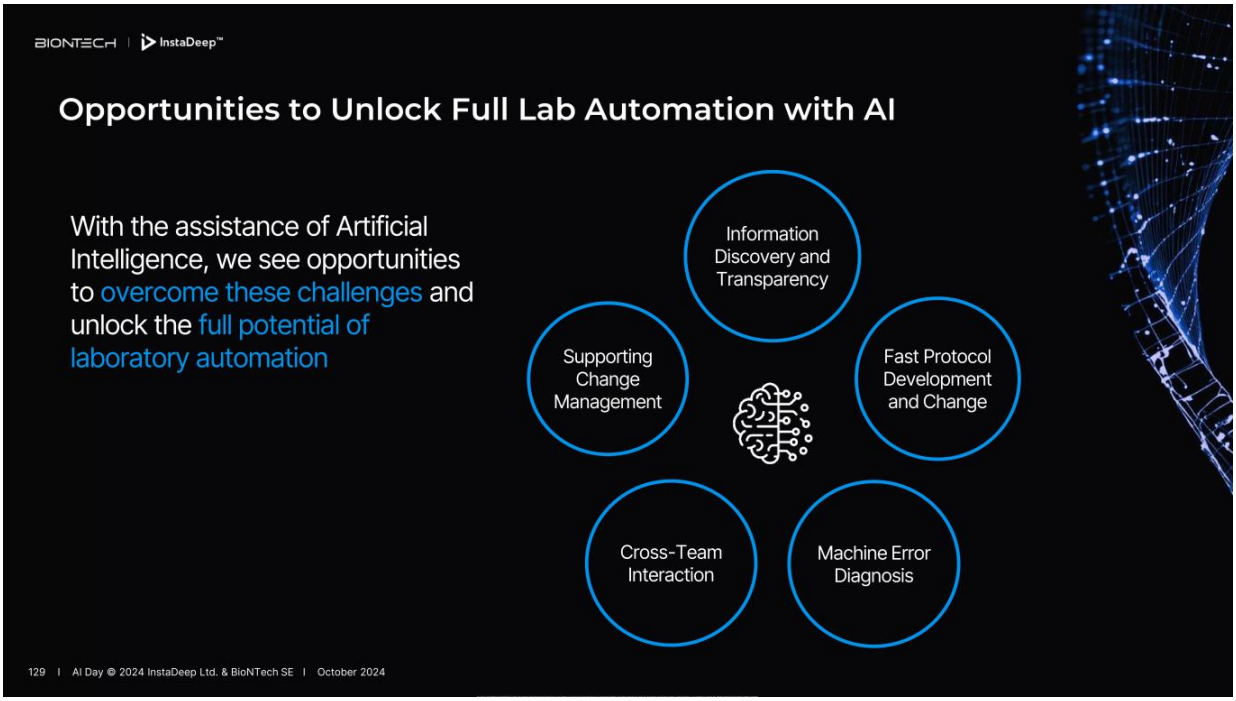


Challenges

Change
R&D is changing constantly

Complexity
High complexity of science and automation

Transparency
Scientists need transparency and trust


BIONTECH | InstaDeep™
Opportunities to Unlock Full Lab Automation with AI
With the assistance of Artificial Intelligence, we see opportunities to overcome these challenges and unlock the full potential of laboratory automation
Information Discovery and Transparency
Supporting Change Management
Fast Protocol Development and Change
Cross-Team Interaction
Machine Error Diagnosis
129 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024

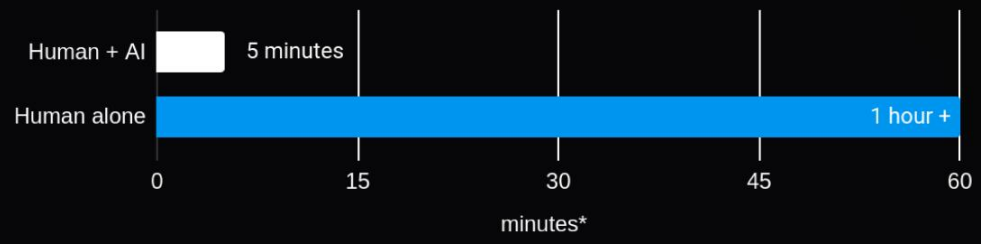
BIONTECH | InstaDeep™

Demo

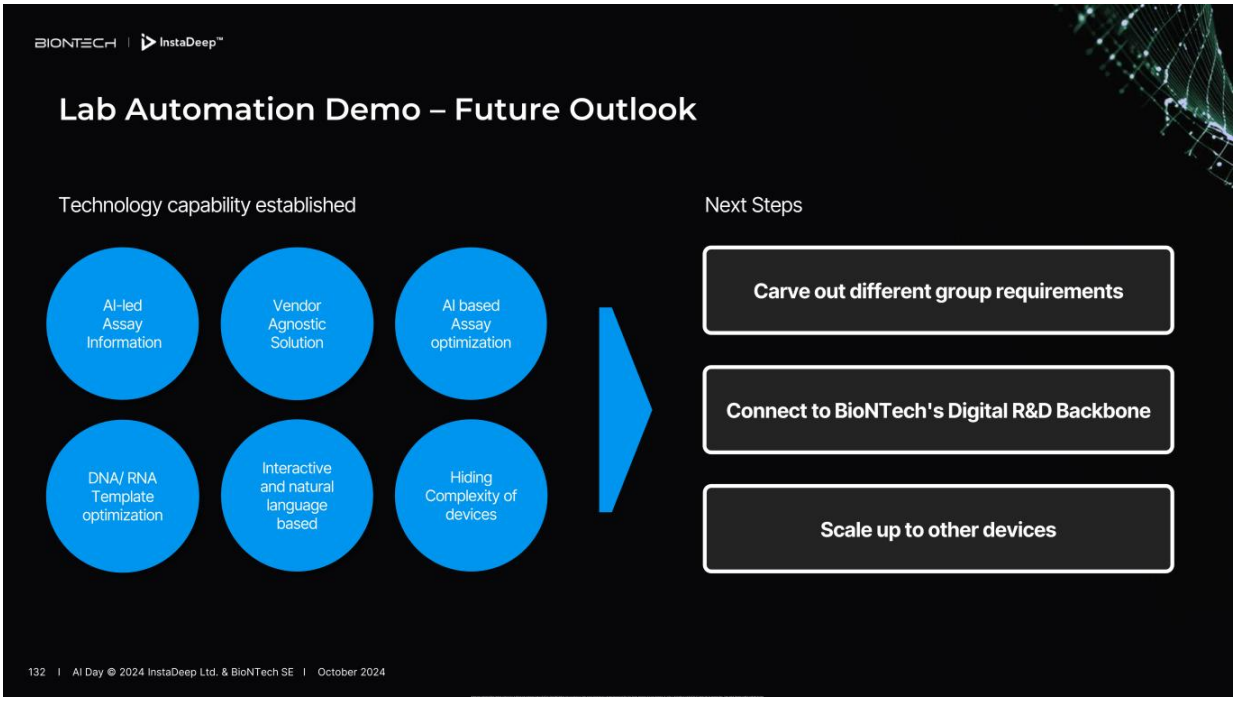
BIONTECH | InstaDeep™

Increased Efficiency Across Laboratory Research Activities

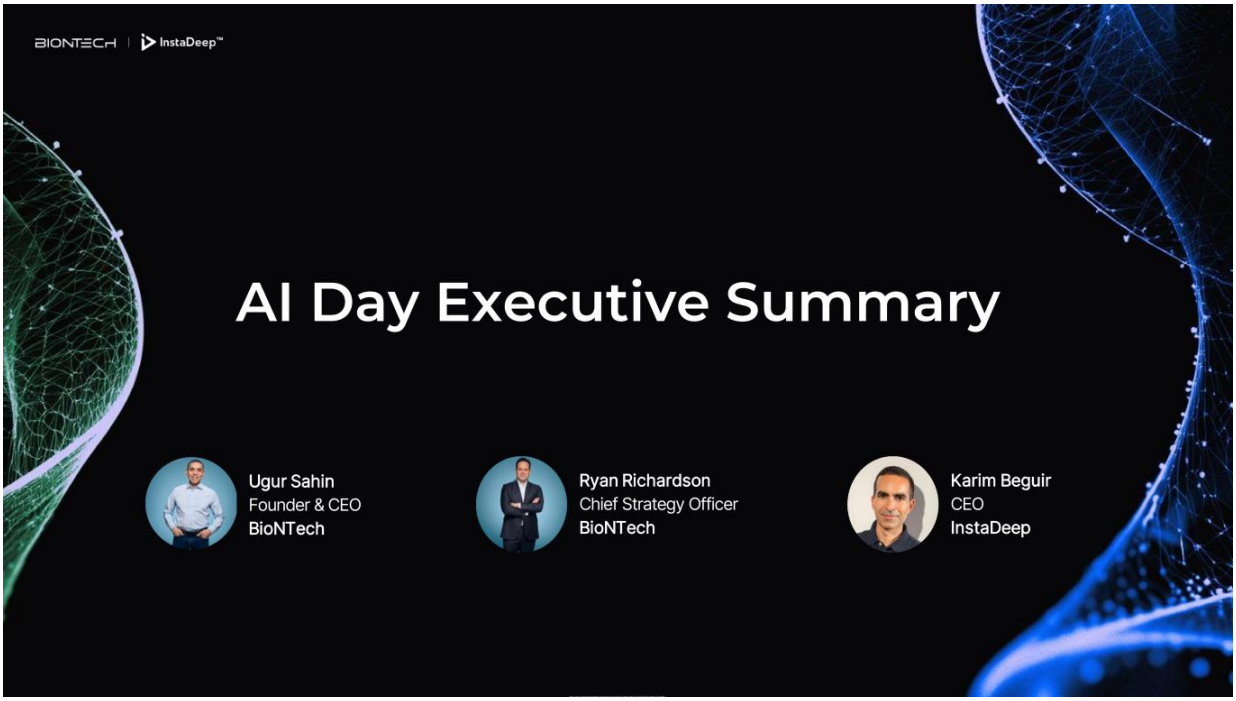
Case Study: liquid handler error diagnosis



*Stated times are for an unexpected error


BIONTECH | InstaDeep™
Lab Automation Demo – Future Outlook
Technology capability established
AI-led Assay Information
Vendor Agnostic Solution
AI based Assay optimization
DNA/ RNA Template optimization
Interactive and natural language based
Hiding Complexity of devices
Next Steps
Carve out different group requirements
Connect to BioNTech's Digital R&D Backbone
Scale up to other devices
132 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
One more thing...
133 | AI Day © 2024 InstaDeep Ltd. & BioNTech SE | October 2024


BIONTECH | InstaDeep™
AI Day Executive Summary
Ugur Sahin
Founder & CEO
BioNTech
Ryan Richardson
Chief Strategy Officer
BioNTech
Karim Beguir
CEO
InstaDeep


BIONTECH | InstaDeep™
Thank You!
THE END