PHARMALINK
CONFERENCE
VIRTUAL • NOVEMBER 15–16, 2022

Analyzing the strengths, opportunities to the weaknesses and threats for AI implementation in biopharmaceuticals

AFDO RAPS HEALTHCARE PRODUCTS COLLABORATIVE

What is Artificial Intelligence?

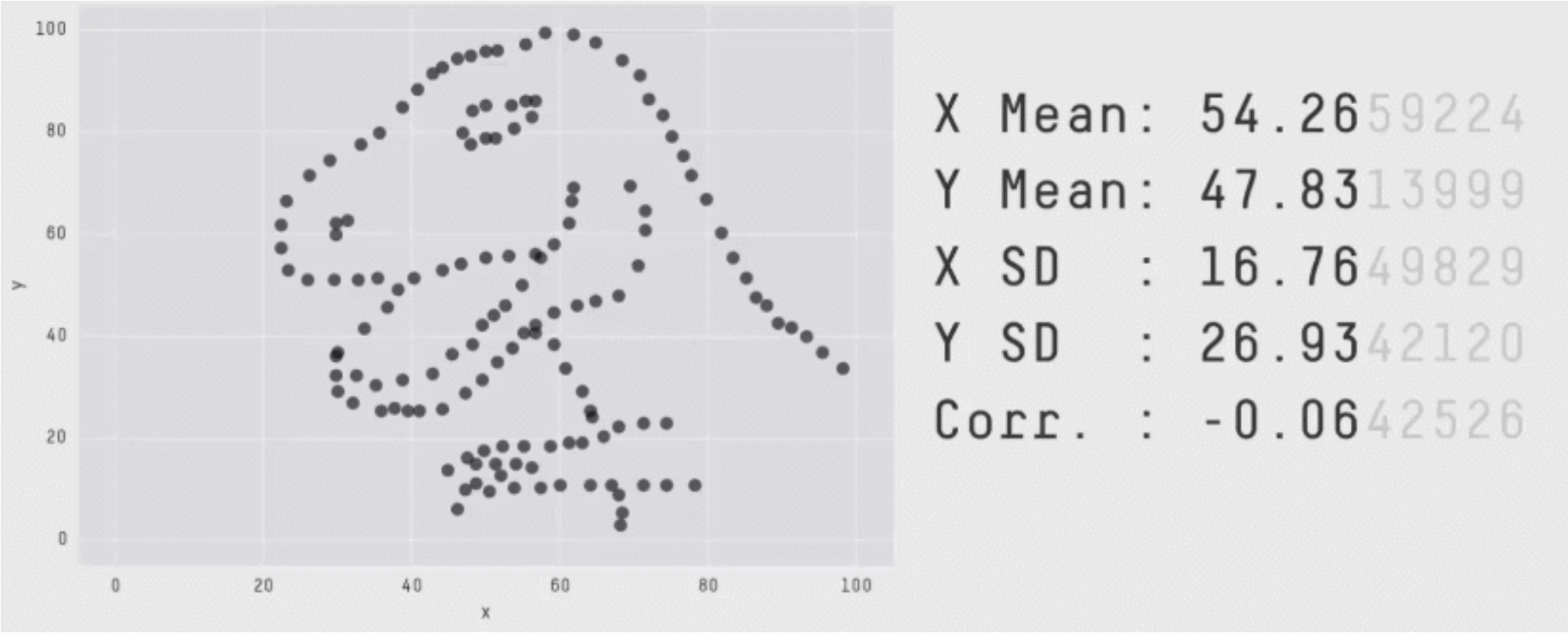
"AI can be thought of as simulating the capacity for abstract, creative, deductive thought - and particularly the ability to learn - using the digital, binary logic of computers."

"Artificial Intelligence (AI) is no longer some bleeding technology that is hyped by its proponents and mistrusted by the mainstream. In the 21st century, AI is not necessarily amazing. Rather, it is often routine. Evidence for the routine and dependable nature of AI technology is everywhere."

"Verification and Validation and Artificial Intelligence," Tim Menzies, Portland State University, Charles Pecheur, NASA Ames Research Center. July 2004.

What Really is AI?

More than Just Multivariable Models...



Mechanistic vs. Probabilistic

- Physics and Engineering mechanics provides the right conditions for the ideal or known scenario
- The new probability (AI models) defines the real conditions without the physical and chemical basis

Classic mechanics

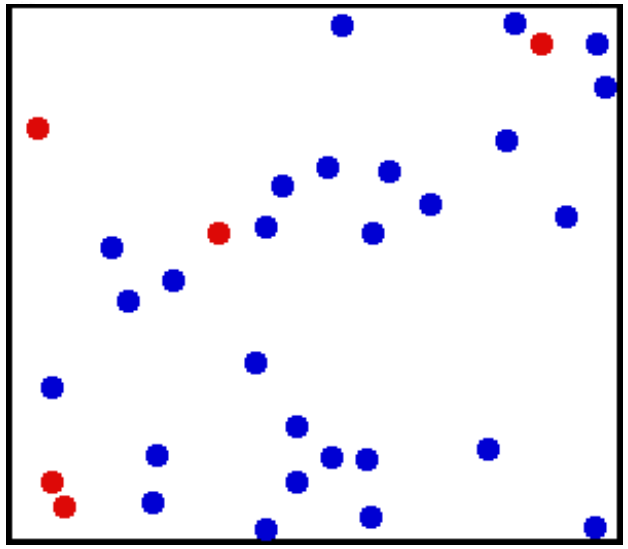
$$\frac{(p_1-p_2)}{p1} < F_\gamma \cdot x_T \rightarrow$$

$$Q_a = \frac{1}{60} \cdot 4.17 \cdot C_v \cdot p_1 \cdot \left(1 - \frac{\frac{p_1-p_2}{p_1}}{(3F_\gamma \cdot x_T)}\right) \cdot \sqrt{\frac{\frac{p_1-p_2}{p_1}}{(T_a+273.15)}}$$

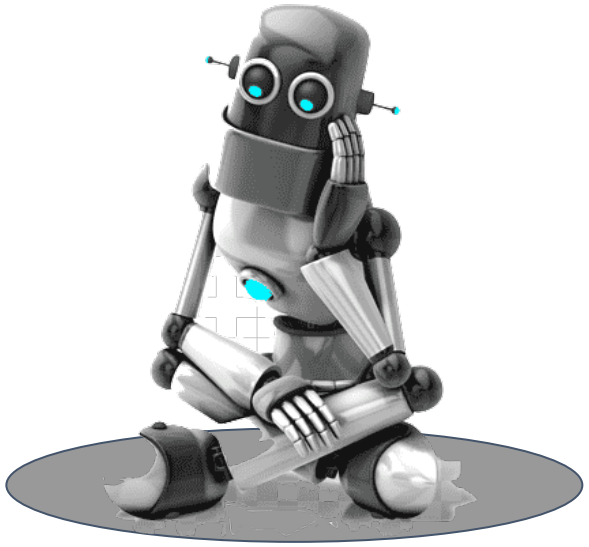
$$\frac{(p_1-p_2)}{p1} \geq F_\gamma \cdot x_T \rightarrow$$

$$Q_a = \frac{1}{60} \cdot 0.667 \cdot 4.17 \cdot C_v \cdot p_1 \cdot \sqrt{\frac{F_\gamma \cdot x_T}{T_a+273.15}}$$

Statistical mechanics



Understanding AI



- **Training/Test data**
  - Problem's dataset
- **Algorithm**
  - Mathematical **procedure** that creates the Model from the training data
- **Model**
  - Mathematical **system** that has been created from the exploration of a data set. It's created after an extensive learning process referred as **training.**
- **Prediction**
  - Single inference over a model with an unseen sample
- **Evaluation**
  - Score evaluation of the Test dataset

Understanding AI: Data (the Secret Sauce)

Data must be prepared before to use it: DS invest the 80% of their time on it

7 steps to consider when preparing data

Findable, **A**ccessible, **I**nteroperable, **R**eusable principle

ALCOA+++

Vas Narasimhan, CEO of Novartis AG, in a 2018

"We've had to spend most of the time just cleaning the data sets before you can even run the algorithm"

Biomanufacturing 4.0: product and process under control

Manufacturing Science

The body of knowledge available for a specific product and process, including critical-to-quality product attributes and process parameters, process capability, manufacturing and process control technologies and quality systems infrastructure.

Source: PhRMA Quality Technical Committee (2003)

PAT

(...) The applicant should demonstrate an enhanced knowledge of product performance over a range of material attributes, manufacturing process options and process parameters

(...) Real-time quality control, leading to a reduction of end-product release testing

(...) A monitoring program (e.g., full product testing at regular intervals) for verifying **multivariate prediction** models

Source: ICH Q8, step 4 (2009)

AI is already here

nature

Explore content | About the journal | Publish with us

nature > news > article

Published: 05 April 2017

Machine learning predicts the look of stem cells

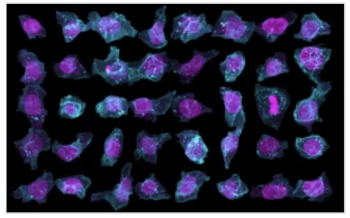
Amy Maxmen

Nature (2017) | Cite this article

610 Accesses | 4 Citations | 524 Altmetric | Metrics

Website contains thousands of 3D stem cell images and could eventually help with better understanding diseases like cancer.

No two stem cells are identical, even if they are genetic clones. This stunning diversity is revealed today in an enormous publicly available online catalogue of 3D stem cell images. The visuals were produced using deep learning analyses and cell lines altered with the gene-editing tool CRISPR. And soon the portal will allow researchers to predict variations in cell layouts that may foreshadow cancer and other diseases.



Structural differences in the DNA (purple) and cellular membrane (blue) of genetically identical stem cells. Credit: Allen Institute for Cell Science

nature

Explore content | About the journal | Publish with us | Subscribe

nature > news > article

NEWS | 22 July 2019

AI protein-folding algorithms solve structures faster than ever

Deep learning makes its mark on protein-structure prediction.

Matthew Hutson

Predicting protein structures from their sequences would aid drug design. Credit: Edward Kinsman/Science Photo Library

nature

Explore content | About the journal | Publish with us | Subscribe

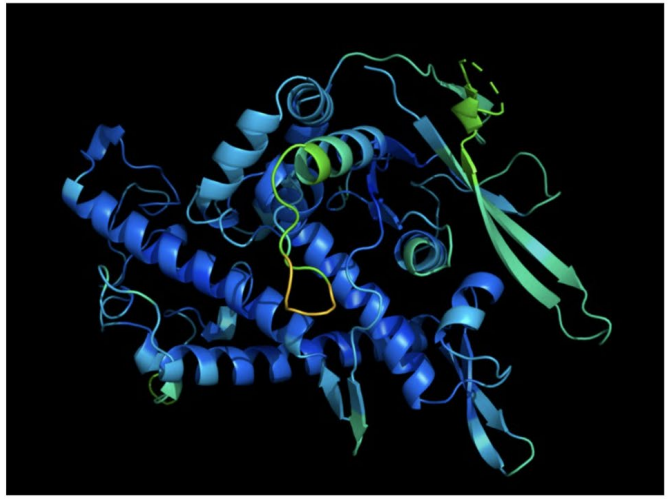
nature > news > article

NEWS | 30 November 2020

'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures

Google's deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

Ewen Callaway



A protein's function is determined by its 3D shape. Credit: DeepMind

AI SUMMIT
COLUMBUS, OH • OCTOBER 25–27, 2022

CPV of the Future

AFDO RAPS HEALTHCARE PRODUCTS COLLABORATIVE

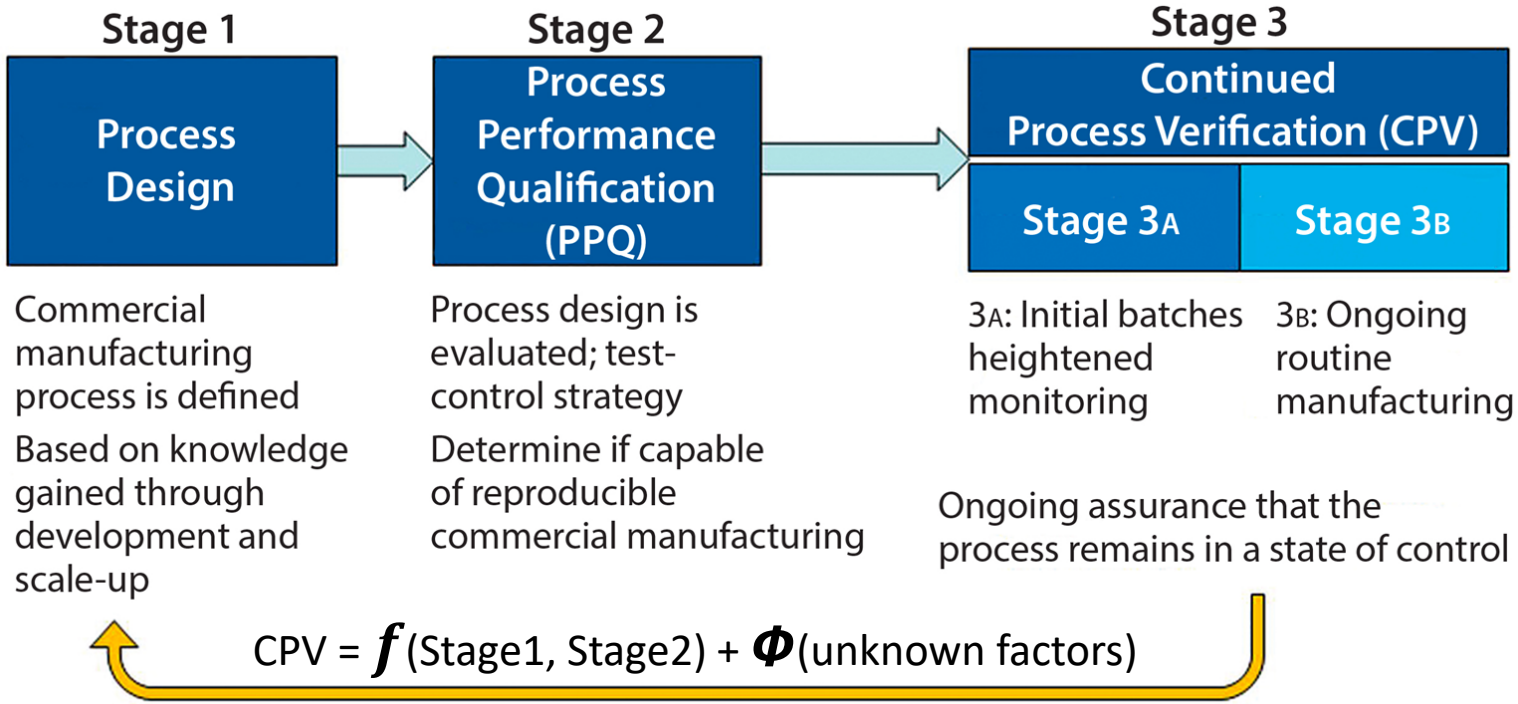
Continued Process Verification (CPV)

Guidance for Industry

Process Validation: General Principles and Practices

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Veterinary Medicine (CVM)

January 2011
Current Good Manufacturing Practices (CGMP)
Revision 1



Real time
- Data acquisition, processing and analysis (IIoT)
- Process verification and control

PHARMALINK CONFERENCE
VIRTUAL • NOVEMBER 15–16, 2022

CPV of the Future

Goal: Create a digital twin to manage a biotech process working under GxP conditions

Q: What does it mean, a digital twin?
Q: What kind of biotech process are we controlling?
Q: How to apply AI within regulatory frameworks?

PDA Interest Group "Process Validation" (EU)

Taskforce 1 | Synthetic data generation to support AI model training
Taskforce 2 | Automated biotech process control
Taskforce 3 | Regulatory considerations (QbD, Data governance, AI Procedures)

PHARMALINK CONFERENCE
VIRTUAL • NOVEMBER 15–16, 2022

Current Team Members

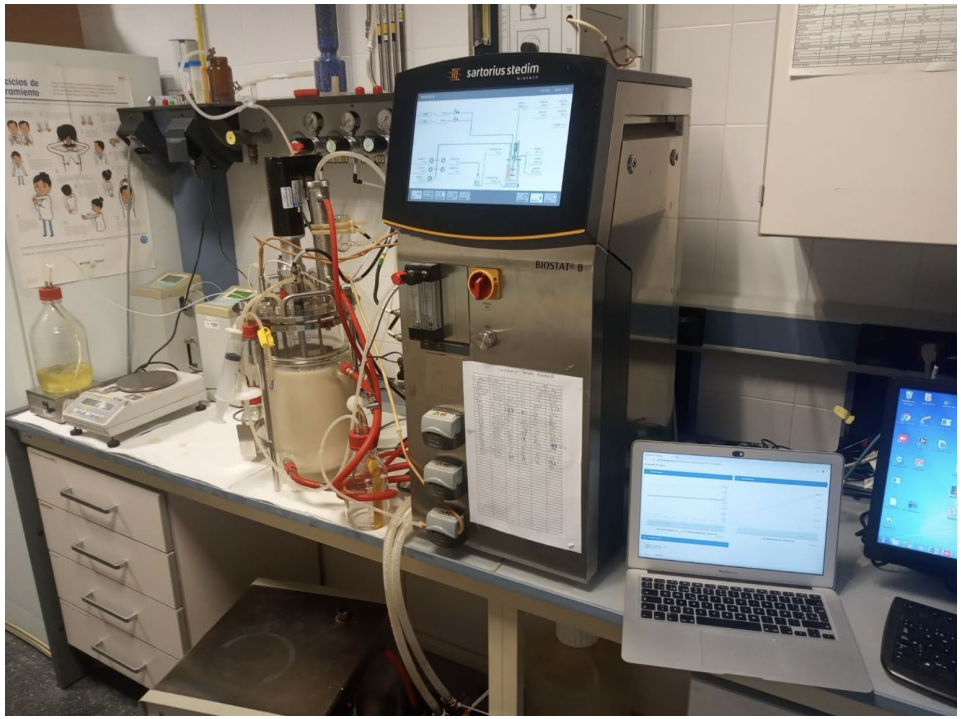
Antonio Moreira *(Execution co-lead*	University of Maryland, Baltimore County	Vice Provost for Academic Affairs
Ben Stevens (PV Team lead)	GSK	Director CMC Policy and Advocacy
Catarina Leitão	4Tune Engineering	CPV Expert
Christophe Agut	Sanofi Pasteur	Head of Process Validation and Statistics Expertise
David Hubmayr (Task Force 1 lead)	CSL Behring	Manager, Process Development & Breakthrough Technologies, R&D
David Lapeña	Infors	Area Sales Manager Southern Europe & Africa
Francisco Valero (Task Force 2 lead)	Universidad Autònoma of Barcelona	Professor and head of department of BioChemical Engineering
Joeri Van Wijngaarden	Aizon	Innovation Lab R&D Project Manager
Mario Stassen (Task Force 3 lead)	AFDO (AI in Operations Team)	BioPharma Regulatory expert

Matt Schmucki	AFDO and AstraZeneca	Lean Coach and CPV Expert
Mauro Giusti (PV Team lead)	Eli Lilly	Director, Technical Services/Mfg Sciences
Nilanjan Banerjee	University of Maryland, Baltimore County	Professor, Computer Science and Electrical Engineering
Sandrine Dessoy	GSK	Science and Technology Innovation Director
Shereya Maiti	Bayer Pharmaceuticals	Senior Scientist
Toni Manzano *(Execution lead)*	AFDO and Aizon	CSO and Co-founder
Ciro Cottini	Chiesi	Digital, Data & Modelling Head
Holger Mueller	Bluesens	Director
Maria A. Batalha	4Tune Engineering	Data Scientist

AFDO RAPS HEALTHCARE PRODUCTS COLLABORATIVE

Bioreactor Fermentation Process

→ Optimise the hypoxic conditions for *Pichia Pastoris* yeast to maximize production
→ Study the effect of specific growth rates (DoE) on yield & protein stability

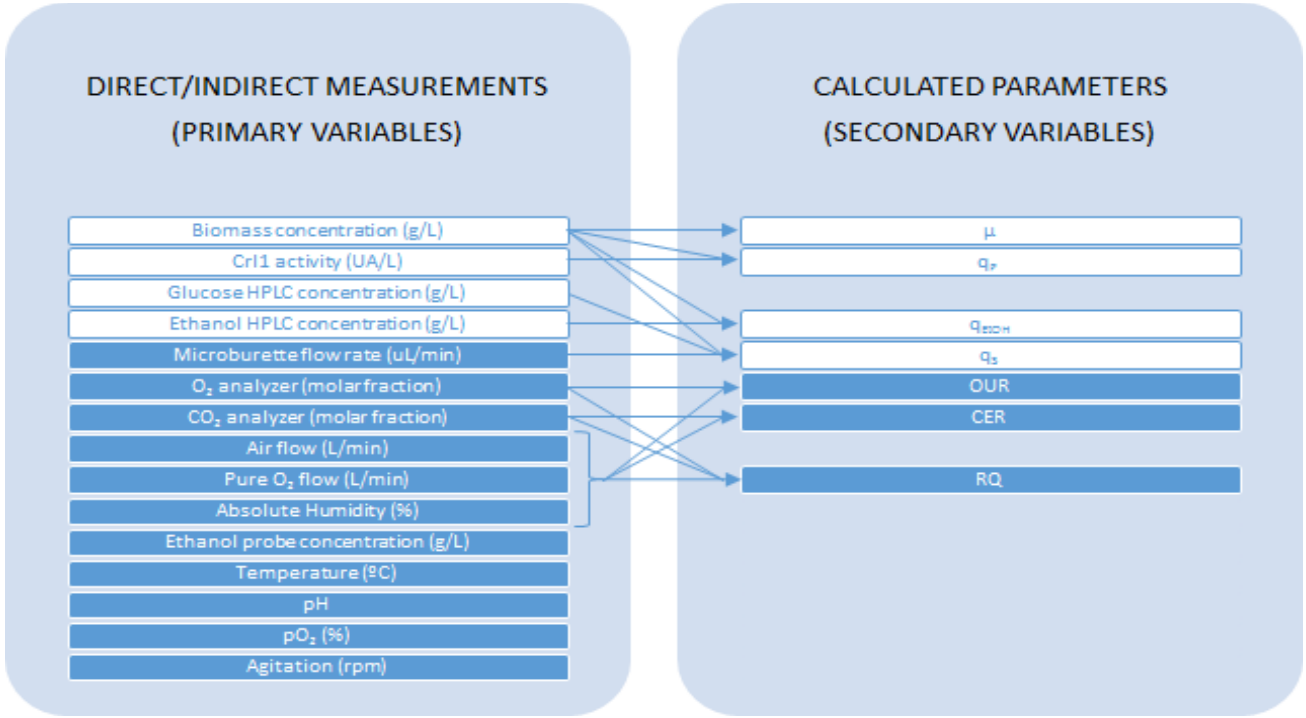


	Normoxic			Hypoxic		
	Good	Average	Bad	Good	Average	Bad
Phase I (Batch)	FBHPX2 FBHPX5 !! FBHPX6 FBHPX9	FBHPX8		FBHPX3 FBHPX4 !! FBHPX10 FBHPX11	FBHPX7	
Phase II (Adaptation)	FBHPX2 FBHPX5 !! FBHPX6 FBHPX8	FBHPX9		FBHPX3 FBHPX4 !! FBHPX7 FBHPX10 FBHPX11		
Phase III (Early Fed Batch)	FBHPX2 FBHPX5 !! FBHPX6 FBHPX8 FBHPX9			FBHPX7	FBHPX4 !! FBHPX10 FBHPX11	FBHPX3
Phase IV (Later Fed Batch)	FBHPX5 !! FBHPX6 FBHPX8 FBHPX9	FBHPX2		FBHPX7 FBHPX10 FBHPX11		FBHPX3 FBHPX4 !!

AI-guided Process Monitoring

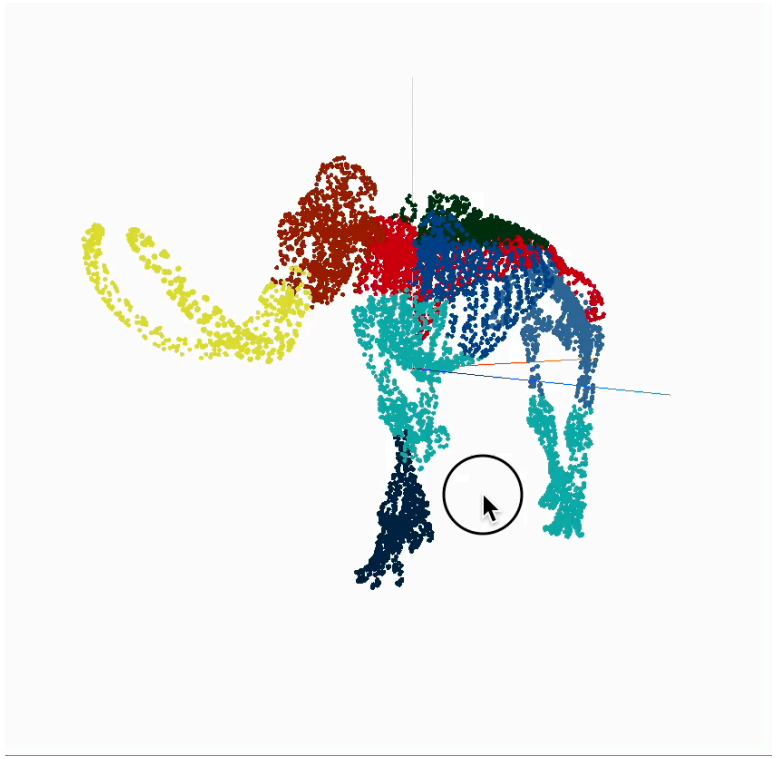
Phase 1: Batch

→ Controlled process, no manual actions
→ Anomalies due to equipment, material qualities or improper behaviour of biomass
→ Multiple relevant factors can be monitored to detect underperforming batches



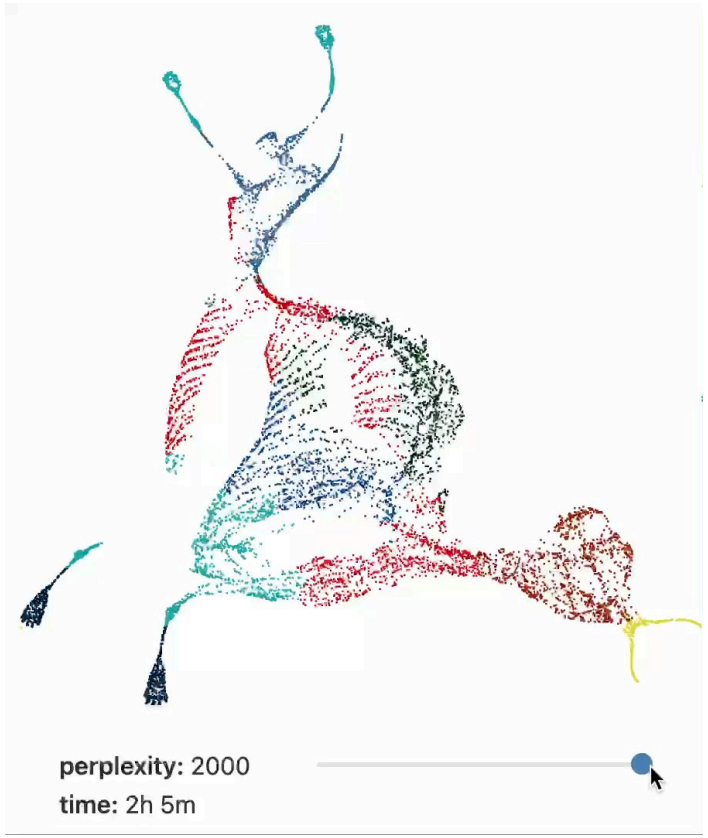
Q: how can we bring value from AI in an automated process without interacting with the unit?

Reducing dimensionality by means of AI



2D UMAP projection

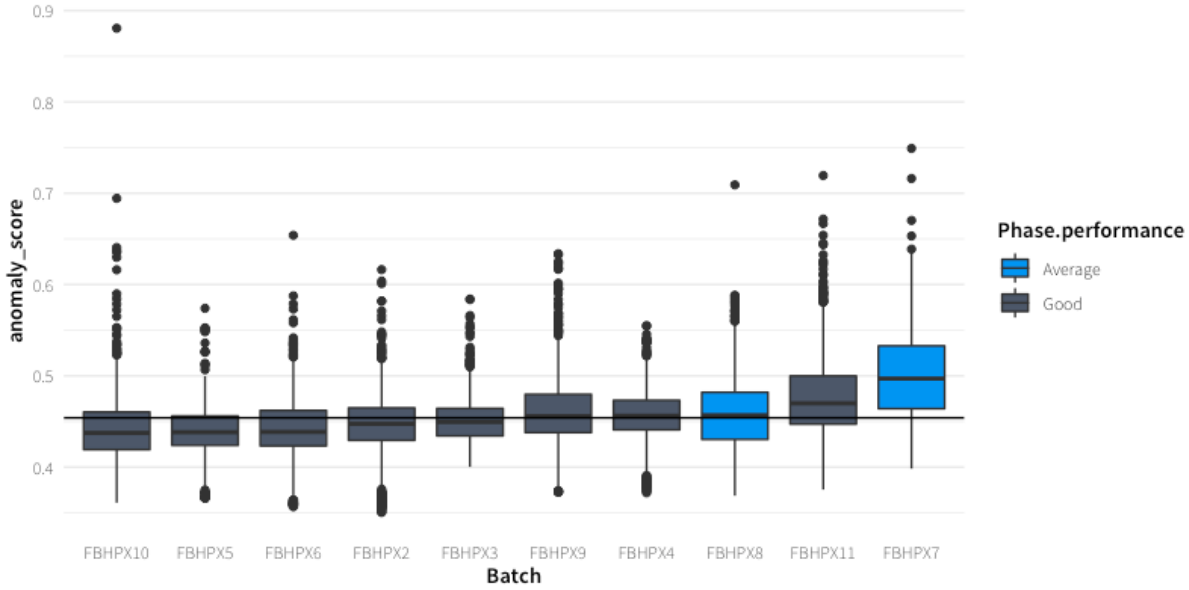
2D t-SNE projection

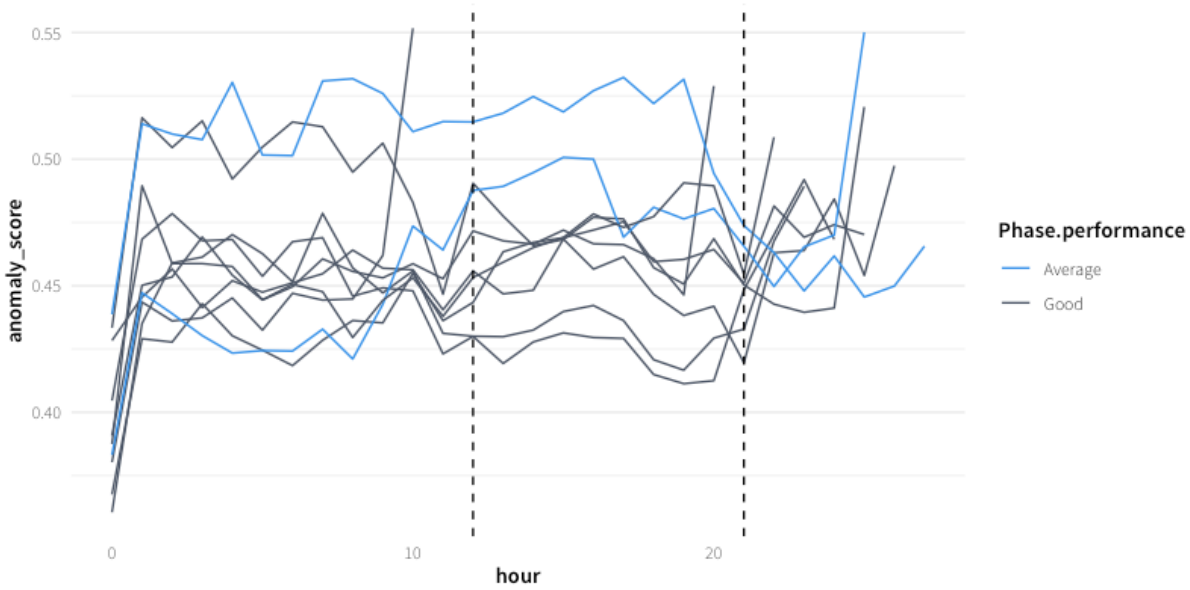


AI-guided Process Monitoring

Phase 1: Batch

→ Controlled process, no manual actions
→ Anomalies due to equipment, material qualities or improper behaviour of biomass
→ Multiple relevant factors can to be monitored to detect underperforming batches

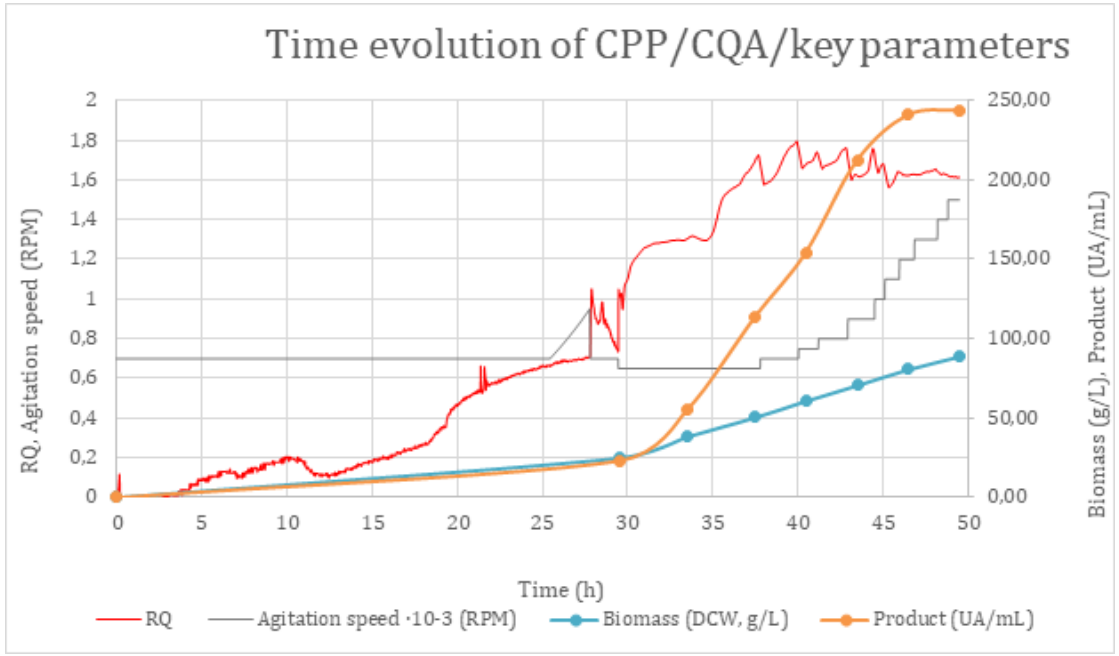


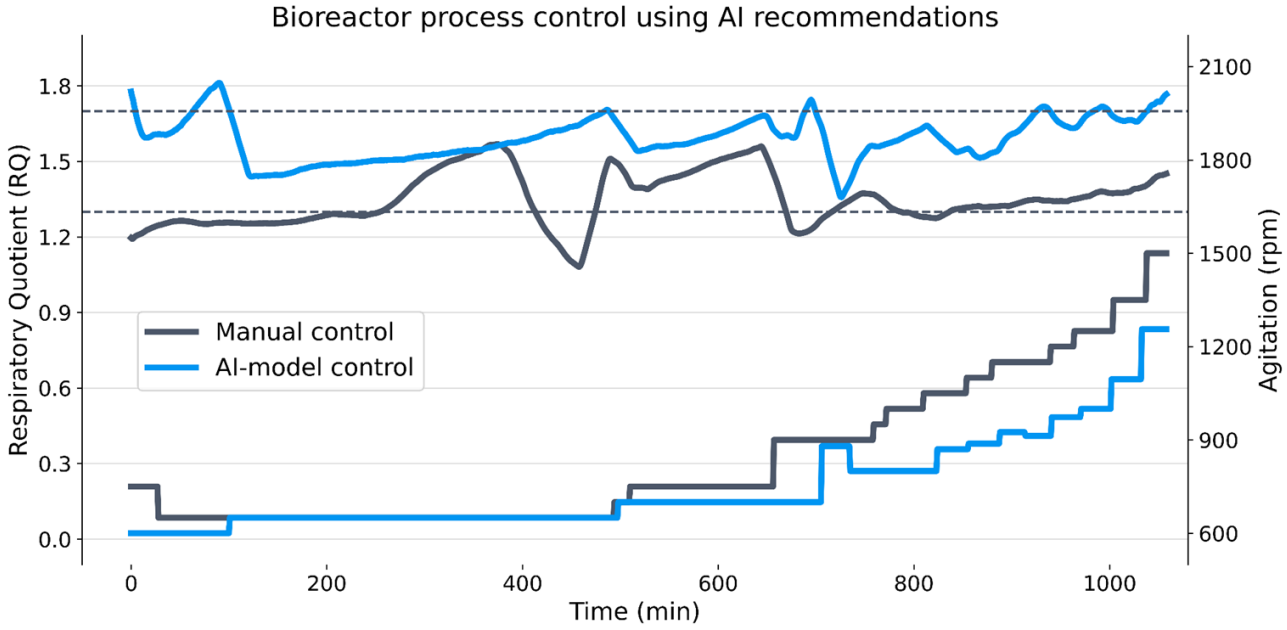


AI-guided Process Control

Phase 2: Fed-Batch

→ Final phase, hypoxic conditions.
→ System requires constant manual control to keep the metabolic parameters within the desired operating range by controlling the Agitation speed.





Source: Gasset et al. (2021) Front. Bioeng. Biotechnol.

Lessons Learned From Phase I

RISK ASSESSMENT

- Important in an initial phase of the project
- Include multidisciplinary team (quality, regulatory, data scientist, CPV expert)

DATA INTEGRITY

- Using cloud-based storage with audit trail to retain the original state of raw data.
- Manual data clean-up should be avoided, as well as insecure data handling and transfer (USB, email etc.). It is better to have raw data available on the platform that handles the AI model life cycle.

PRIOR KNOWLEDGE

- Key enabler to efficiently design and run the DoE.
- Design of DoE, factor selection & operating ranges were defined based on prior knowledge (20 years experience).
- Check equipment responsible for critical measurements before running experiments (ex: In the case of hypoxic fermentations, gas analysis system is crucial, so lab team re-calibrated O2 and CO2 analyzers in each fermentation, checked inlet gas composition periodically, considered gas humidity, etc.)

PHARMALINK CONFERENCE
VIRTUAL • NOVEMBER 15–16, 2022

SWOT Analysis: AI for CPV

STRENGTHS

- Capacity to deal with multivariate and complex reality
- Capacity to deal with the dynamic nature of bioprocesses
- Easy detection of anomalous batches in historical data plus assess upcoming batches in real-time
- Estimating optimal operating conditions for bioprocess unit, including bioprocess efficiency and product quality

WEAKNESSES

- High volumes of data needed to feed AI models
- Predictions are sensitive to “bad quality” data, which leads to rapid deterioration of performance or incorrect conclusions.
- Data management and preparation is time-consuming

OPPORTUNITIES

- CPV highly recommends bioprocess automation, use of PAT, risk assessment, and a deep knowledge of the biomanufacturing process quality attributes. Existing methodologies can be complemented with AI.
- Continued Improvement is always a “work-in-progress” task since the AI model is continuously learning from the data that is being fed.

THREATS

- Impression that AI applications are a “black box”, while regulatory requirements require transparency and regulators need to understand how and why a result came about.
- AI applications can produce valid conclusions that are counter intuitive to those which individuals or even teams of experts derive.
- Deal with model changes over time in a regulatory point of view (freeze the model vs multiple submissions)

PHARMALINK CONFERENCE
VIRTUAL • NOVEMBER 15–16, 2022

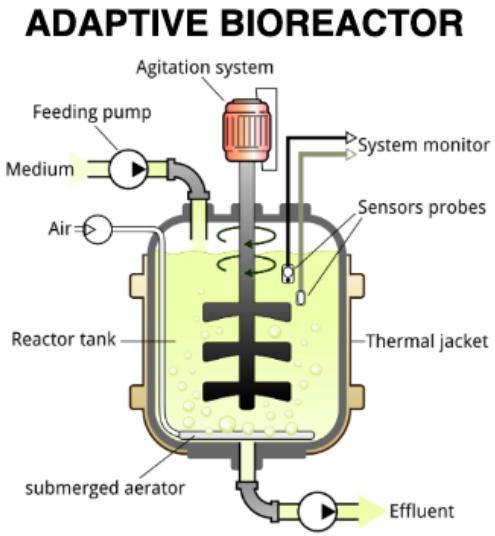
Control Automation For Optimal Data Governance

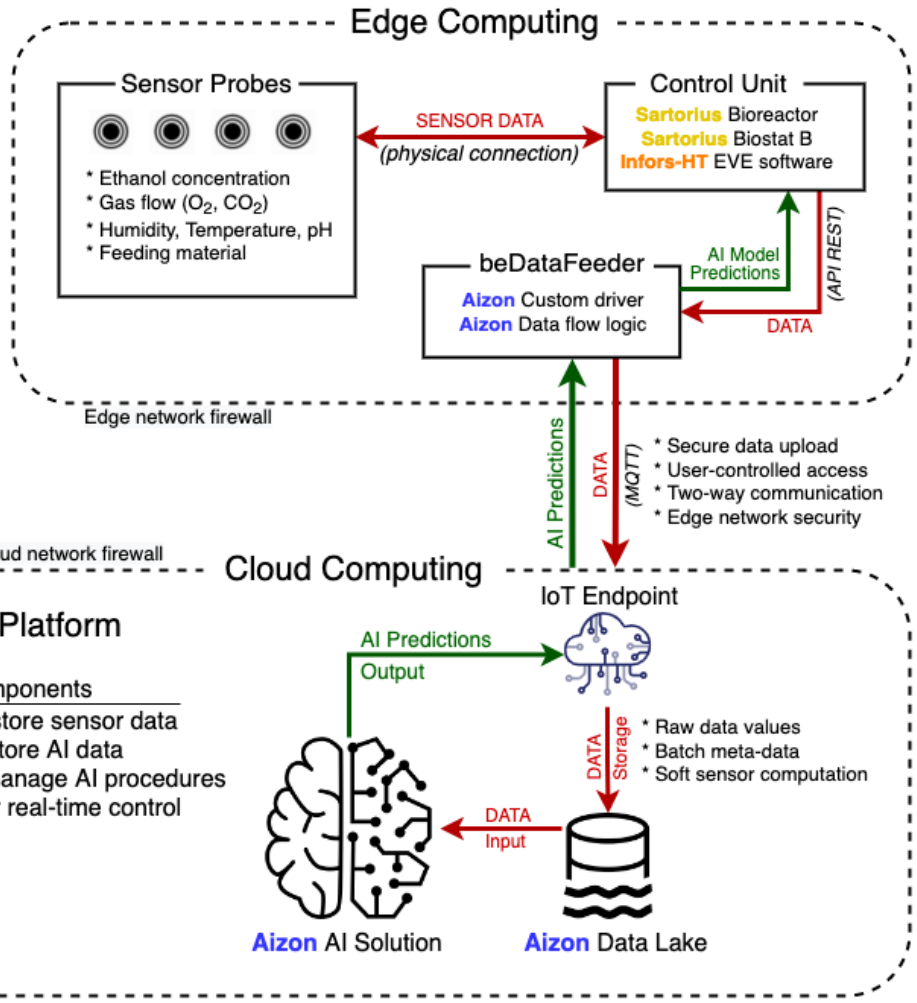
Control Strategy

- PAT & IIoT Technology
- Combination of edge & cloud
- Fully automated data pipeline
- Coverage of full AI lifecycle (train, productivise, monitor)
- Operate in near real-time

Control Parameters

- Storage of 17 raw data variables
- Critical read-out: *respiratory quotient (RQ)*
- Critical control: *agitator speed (AS)*





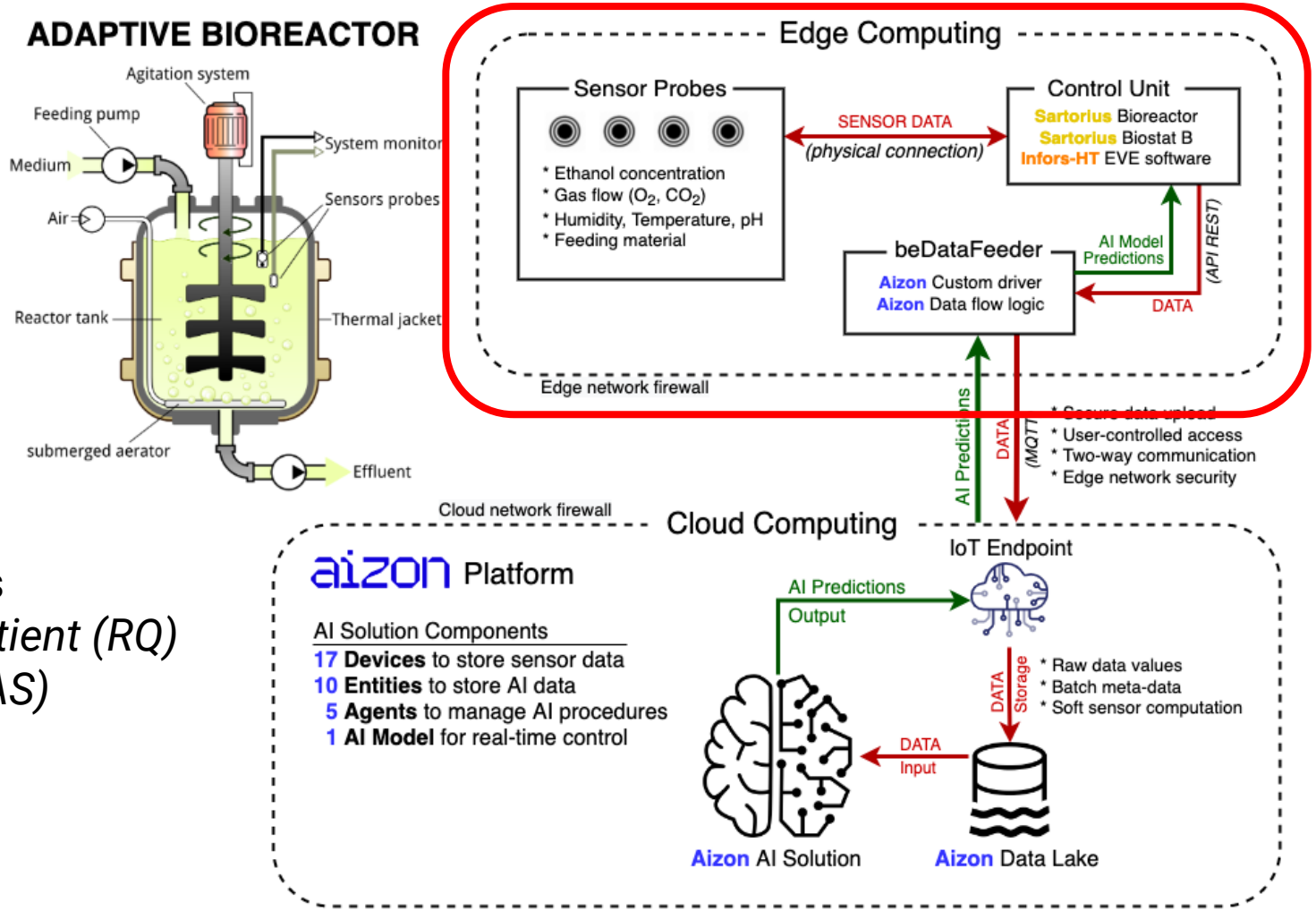
Control Automation For Optimal Data Governance

Control Strategy

- PAT & IIoT Technology
- Combination of edge & cloud
- Fully automated data pipeline
- Coverage of full AI lifecycle (train, productivise, monitor)
- Operate in near real-time

Control Parameters

- Storage of 17 raw data variables
- Critical read-out: *respiratory quotient (RQ)*
- Critical control: *agitator speed (AS)*



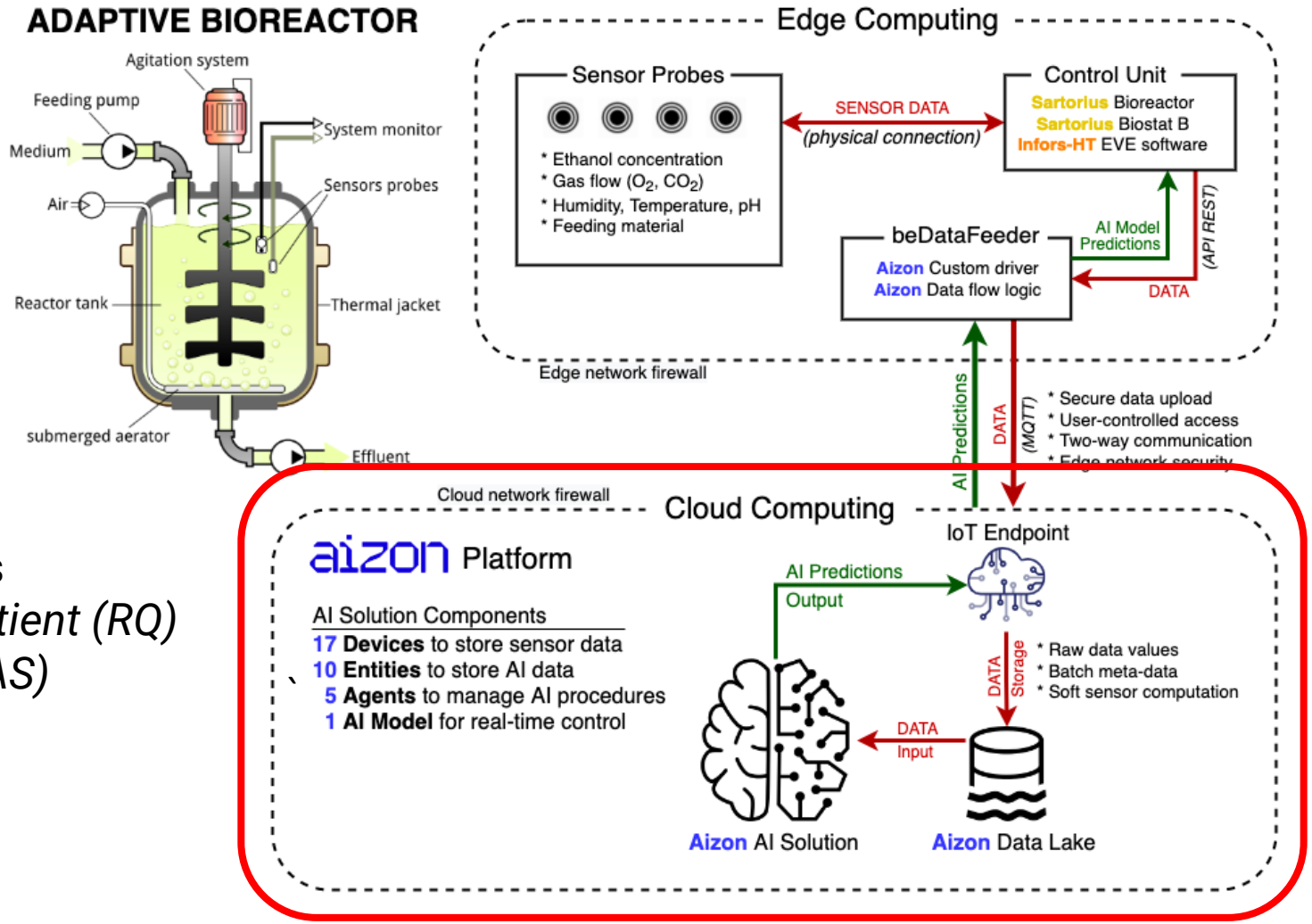
Control Automation For Optimal Data Governance

Control Strategy

- PAT & IIoT Technology
- Combination of edge & cloud
- Fully automated data pipeline
- Coverage of full AI lifecycle (train, productivise, monitor)
- Operate in near real-time

Control Parameters

- Storage of 17 raw data variables
- Critical read-out: *respiratory quotient (RQ)*
- Critical control: *agitator speed (AS)*



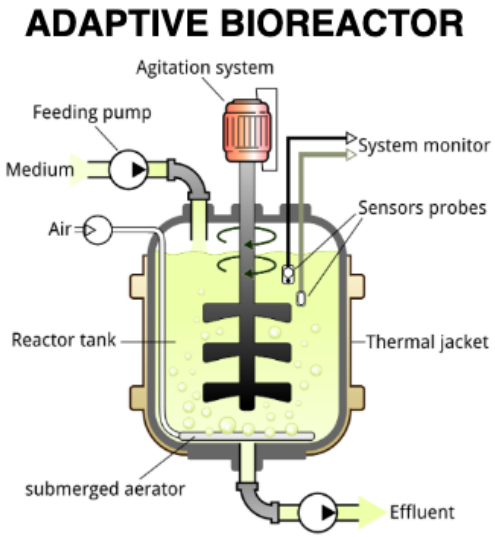
Control Automation For Optimal Data Governance

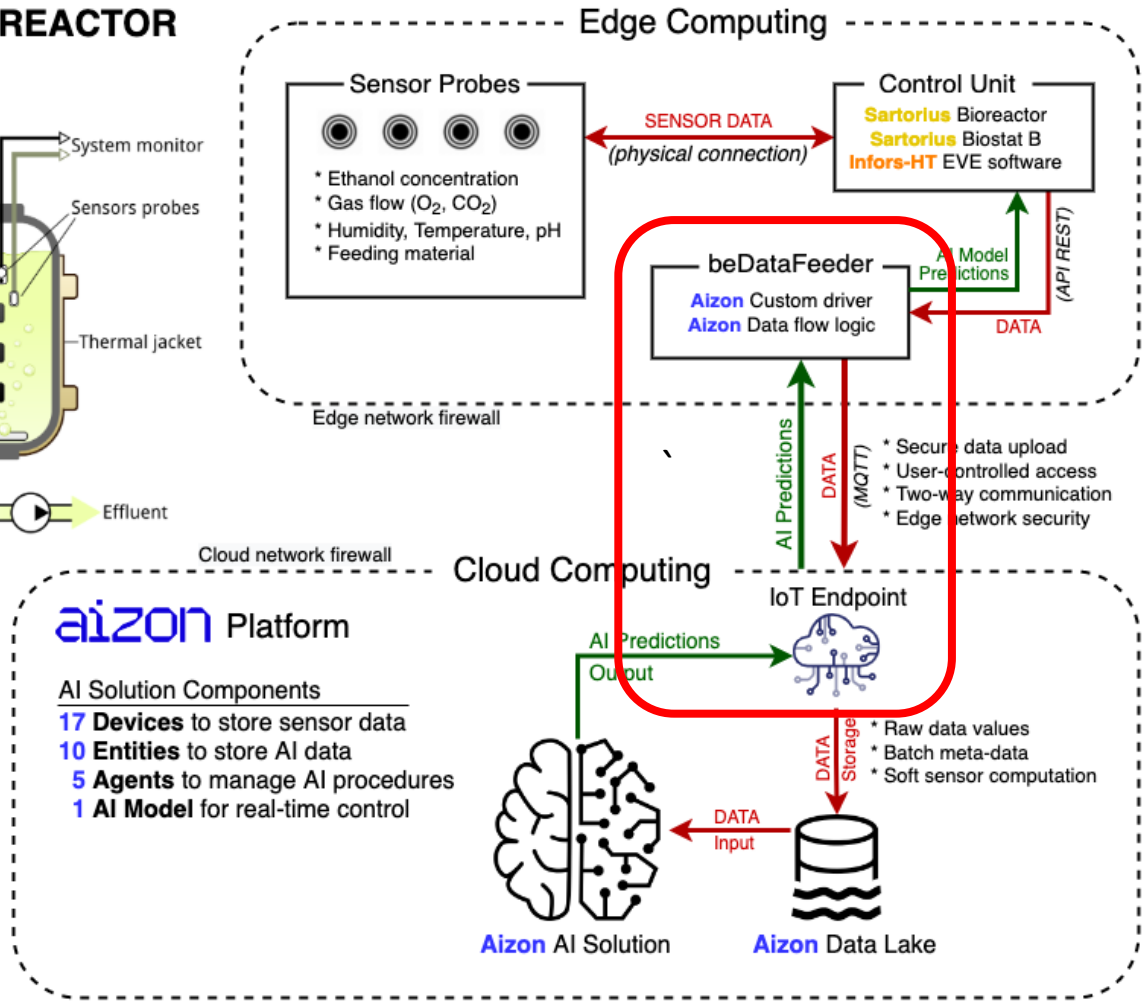
Control Strategy

- PAT & IIoT Technology
- Combination of edge & cloud
- Fully automated data pipeline
- Coverage of full AI lifecycle (train, productivise, monitor)
- Operate in near real-time

Control Parameters

- Storage of 17 raw data variables
- Critical read-out: *respiratory quotient (RQ)*
- Critical control: *agitator speed (AS)*





PHARMALINK CONFERENCE
VIRTUAL • NOVEMBER 15–16, 2022

Current Steps

PDA JPST

PDA Journal of Pharmaceutical Science and Technology

search

Advanced Search

Home | Content | About PDA JPST | Author & Reviewer Resources | JPST Access and Subscriptions | Support

Other | Research

CPV of the Future: AI-powered continued process verification for bioreactor processes

Andrej Ondracka, Arnau Gasset, Xavier García-Ortega, David Hubmayr, Joeri B.G. van Wijngaarden, José Luis Montesinos-Seguí, Francisco Valero and Toni Manzano

PDA Journal of Pharmaceutical Science and Technology September 2022, pdajpst.2021.012665; DOI: https://doi.org/10.5731/pdajpst.2021.012665

Article | References | Info & Metrics | PDF

Abstract

According to the standard guidelines by the FDA, process validation in biopharma manufacturing encompasses a lifecycle consisting of three stages: Process design (PD), Process qualification (PQ), and Continued process verification (CPV). The validity and efficiency of the analytics methods employed during the CPV require extensive knowledge of the process. However, for new processes and new drugs, such knowledge is often not available from PPQV. In this work, the suitability of methods based on machine learning/artificial intelligence (ML/AI) for the CPV applied in bioprocess monitoring and cell physiological control of the yeast Pichia pastoris (Komagataella

In This Issue

PDA Journal of Pharmaceutical Science and Technology
Vol. 76, Issue 5
September/October 2022
Table of Contents
Index by Author
Complete Issue (PDF)

Generierung synthetischer Batch-Daten durch künstliche Intelligenz

David Hubmayr[1], Nilanjan Banerjee[2], Joeri van Wijngaarden[3], Toni Manzano[3]

[1]CSL Behring AG, Bern, Schweiz [2]Fakultät für Informatik und Elektrotechnik. Universität Maryland, Baltimore County, Baltimore, Maryland, USA [3]Aizon Inc., Barcelona, Spanien

Korrespondenz: DI David Hubmayr, Wankdorfstr. 10, 3014 Bern (Schweiz), E-Mail: david.hubmayr@cslbehring.com

ZUSAMMENFASSUNG

In diesem Beitrag wird ein flexibler Ansatz zur Erzeugung synthetischer Batch-Daten, die sich aus multivariaten Zeitreihen zusammensetzen, vorgestellt. Einer der am meisten übersehenen Einflussfaktoren in Bezug auf Künstliche Intelligenz (KI) ist eine umfassende und qualitativ hochwertige Datenbasis. Oft steht diese nur limitiert zur Verfügung. Synthetisch erzeugte Daten können diese Lücke schließen. Im Gegensatz zu Dummy-Daten, erzeugt als Ergänzung zu real gemessenen Daten, bieten *in silico* erstellte synthetische Daten ein hohes Maß an Realismus. Gemäß der Definition handelt es sich bei Dummy-Daten um Ersatzdaten (Musterdaten), die nach dem Zufallsprinzip erzeugt werden. Hierbei werden keine Merkmale des zugrunde liegenden Prozesses und der gemessenen realen Daten berücksichtigt. Der entwickelte KI-Algorithmus ist in der Lage, sowohl realitätsnahe Datenbatches als auch Datenbatches mit einer kontrollierten Streuung der Daten zu generieren, dies erweitert das Feld der möglichen Anwendungsfälle. Die synthetisch erzeugten Datenkurven unterscheiden sich wie geplant innerhalb des Raums, der durch den realen Datensatz aufgespannt wird, zufallsbedingt voneinander. Das Erzielen qualitativ hochwertiger synthetischer Datensätze unter Bereitstellung limitierter realer Datensätze ist ein starker Türöffner für KI-basierte Algorithmen. Synthetisch generierte Daten tragen wesentlich dazu bei, den Einsatz von KI in der pharmazeutischen Herstellung zu verankern und zu beschleunigen, indem sie als datenschutzsicherer Ersatz für reale Daten dienen. Synthetische Daten unterliegen nicht den Datenschutzbestimmungen und überwinden das Risiko der Re-Identifizierung.

KEYWORDS

- Design of Experiments
- Data Science
- Künstliche Intelligenz
- Synthetische Daten
- Synthetische Batches
- Bioprozess

Pharm. Ind. 84, Nr. 2, 264–270 (2022)

ABSTRACT

Generation of synthetic batch data through artificial intelligence

In this paper, a flexible approach to generating synthetic data batches, comprised of multivariate time-series synthetic datasets, is presented. One of the most overlooked influential factors of modern Artificial Intelligence (AI) approaches is an ample and high-quality database. Quite often, ample, and high-quality data is only available to a limited extend. Synthetically generated data can close this gap. Unlike dummy data, in-silico created synthetic data gives unprecedented levels of realism. As per definition, dummy data is mock data generated at random as a substitute for real data in testing environments. In contrast to the simple generation of random substitute data, this effort presents the creation of synthetic data for in-silico generation of additional batches, considering the characteristics of the underlying process and measured real data. Both aspects for synthetic data generation, quality, and quantity of data, are lined out and verified. Inherent to the synthetic data is its ability to not only generate realistic synthetic batch data but also to generate batches with a controlled spread in data if required, broadening the field of potential use cases. As planned, the synthetically generated data curves differ from each other randomly within the space spanned by the real data set. Achieving high-quality synthetic datasets while providing limited real-world datasets is a strong door opener for AI-based algorithms. Synthetically generated data significantly contributes to rooting and accelerating the use of AI in pharma by working as a privacy-secure drop-in replacement for real data. Synthetic data is exempt from privacy regulations and overcomes data re-identification risks.

Einleitung

Der Ansatz zur Entwicklung von Arzneimitteln und ihr jeweiliger Herstellungsprozess variieren von Produkt zu Produkt und von Unternehmen zu Unternehmen, wobei entweder ein empirischer Ansatz oder ein systematischerer Ansatz – als Quality by Design (QbD) bezeichnet – oder eine Kombination aus beiden verfolgt wird,

Conclusions and Next Steps

1. Multivariate analytics methods, based on machine learning & AI, can be a valuable tool for both monitoring and control of biopharma manufacturing bioprocesses to help improve its efficiency and to assure product quality.

2. Initial phases of the project focused mostly on overcoming technological challenges related to cyber security, data integrity and cross-system communication. The next steps will focus on improvements to system validation and reproducibility of data & AI models, following GMP standards.

3. Regulatory considerations require us to redefine the conditions in which we can develop and industrialise AI for manufacturing. This is especially relevant for data integrity, risk assessment, AI lifecycle management.

PHARMALINK CONFERENCE
VIRTUAL • NOVEMBER 15–16, 2022

Acknowledgements

Taskforce 1

David Hubmayr (CSL Behring)
Nilanjan Banerjee (UMBC)
Joeri Van Wijngaarden (Aizon)

Taskforce 2

Francisco Valero & UAB (UAB)
Joeri Van Wijngaarden (Aizon)
Toni Manzano (Aizon)
David Lapeña (Infors)
Ciro Cottini (Chiesi)

Taskforce 3

Mario Stassen (AFDO)
Matt Schmucki (AZ)
Catarina Leitão (4Tune)
Antonio Moreira (UMBC)
Agnes Hardy-Boyer (Sanofi)
Ben Stevens (GSK)
Sandrine Dessoy (GSK)
Maria Batalha (4Tune)
Holger Mueller (Bluesens)

PQRI Product Quality Research Institute
Development | Product Quality | Biopharmaceutics

AFDO CONNECT • SHARE • IMPACT • PROTECT

PDA Parenteral Drug Association

AFDO RAPS HEALTHCARE PRODUCTS COLLABORATIVE

AFDO RAPS

PHARMALINK CONFERENCE
VIRTUAL • NOVEMBER 15–16, 2022

Thank you!

AFDO RAPS HEALTHCARE PRODUCTS COLLABORATIVE

PHARMALINK CONFERENCE
VIRTUAL • NOVEMBER 15–16, 2022

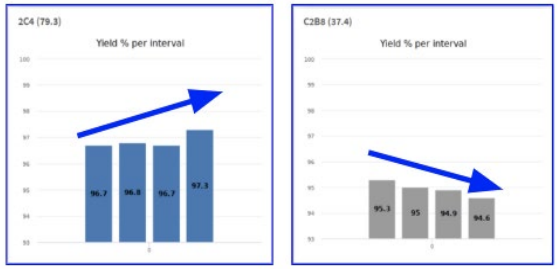
Summary of case studies

GRIFOLS

- **400M** key data values digitized
- **55%** variance explained by AI
- **4%** yield increase
- **$4.1M** savings COGS initially
- **$11.6M** savings COGS by end of 2022

Genentech

- **277g** loss of mAB per batch prevented
- Potential revenue gain of up to **$680M** per year in product
- **$19-78M** savings COGS by end of 2022

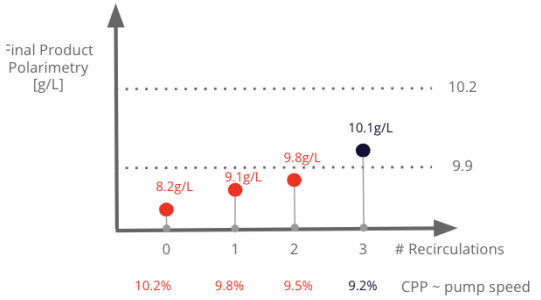


novo nordisk®

- **3h FTE/Day** reduction in time used for performance board reporting
- **11%** OEE stability increase
- **Reactive to Proactive** prediction
- **Manual Reporting** eliminated
- **Faster & more accurate** planning for line personnel and technicians

FERRING PHARMACEUTICALS

- **61%** reduction in necessary runs
- **Prevented** all batch-related data loss
- **100%** Right First Time batches



AFDO RAPS HEALTHCARE PRODUCTS COLLABORATIVE