Project Title: **AIQSP – Generative AI-Driven Quantitative Systems Pharmacology for Enhanced Drug Development**

Call: HORIZON-HLTH-2025-01-TOOL-03: Leveraging.multimodal.data.to.advance. Generative.Artificial.Intelligence.applicability.in.biomedical.research.(GenAI0EU)

Coordinator: Prof. Igor Goryanin, UoE, UK

Participants: Dr Oleg Demin, InSysBio, Cyprus

Who else? To discuss

Project Type: Research and Innovation Action (RIA)
Duration: 36 months To discuss

Project Summary

AIQSP aims to revolutionize drug development by integrating **Quantitative Systems Pharmacology (QSP)** with **Generative AI** to form a robust, intelligent platform for predictive modeling and personalized medicine.

QSP provides a mechanistic framework for simulating drug interactions within biological systems by combining computational models with diverse experimental and clinical datasets (in vitro, in vivo, clinical trials). The introduction of **Virtual Patient Populations (VPpop)** further enhances this by predicting heterogeneity in drug response across subgroups.

However, QSP modeling remains underutilized due to its complexity and the need for specialized expertise. To address this, **AIQSP will develop "QSP-GPT"**, a generative AI tool trained on QSP simulations that makes powerful predictive modeling accessible to non-experts in drug development, such as pharmacologists and clinicians.

The platform will also include:

- **Automated integration of multimodal data**, including genomic, transcriptomic, proteomic, and kinetic datasets. To discuss
- Continuous **tracking and incorporation of new public-domain data** into QSP models using AI for real-time model evolution.
- Integration with tools like **AlphaFold** to strengthen mechanistic understanding and protein-target predictions.
- Application to **breast cancer**, mental disorders, with model predictions validated in cell lines, animal models, and pilot clinical trials. To discuss.

Key Innovations

- **Generative AI-assisted QSP** for interpretability, usability, and real-time learning.
- **Automated literature and data mining** for model updates and expansion.
- **User-friendly interface** enabling cross-functional teams to leverage advanced modeling.
- **Validated AI predictions** on real-world biomedical data and clinical endpoints. To discuss

---

Looking for Partners With Expertise In

- AI/ML and large language model development
- Bioinformatics and systems biology
- Drug development and clinical trial design
- Regulatory sciences and health data standards
- Software platforms in biomedicine

Targeted Partners: AI developers, QSP modelers, bioinformaticians, pharma SMEs, regulatory scientists, clinical researchers, and software engineers.

Scientific AI Component of the AIQSP Proposal

1. Introduction

The AIQSP project seeks to transform biomedical research and drug development through the fusion of Quantitative Systems Pharmacology (QSP) and cutting-edge Artificial Intelligence (AI). This extension aims to elaborate on the scientific direction of the AI component, focusing on the development of next-generation generative large language models (GLLMs), deep neural networks (DNNs), and their integration with biomedical simulations and multimodal datasets.

2. Generative Large Language Models (GLLMs) for Biomedical Simulation

GLLMs, such as GPT-4 and its successors, have demonstrated remarkable abilities in natural language understanding, knowledge extraction, and reasoning. The AIQSP project proposes to develop domain-specific GLLMs tailored for biomedical applications. These models will be trained and fine-tuned on:

- Biomedical literature (e.g., PubMed, PMC)
- Clinical trial reports and regulatory filings (e.g., EMA, FDA databases)
- Drug-target interaction datasets (e.g., ChEMBL, DrugBank)
- Biological pathway and disease mechanism ontologies (e.g., Reactome, KEGG, Gene Ontology, EHMN, Transfac)

A specialized GLLM, termed **QSP-GPT**, will be trained using a corpus of QSP model descriptions, simulation results, and mechanistic annotations. Its primary functions will include:

- Automatic generation of hypotheses for mechanistic modeling
- Interpretation and summarization of simulation outcomes
- Suggestion of model refinements and parameter estimations
- Assistance in regulatory and clinical documentation based on model predictions

These capabilities will be enhanced with retrieval-augmented generation (RAG) mechanisms that dynamically link simulation results and structured databases during inference.

3. Deep Neural Networks for Multimodal Integration

Deep Neural Networks (DNNs) will be used to integrate and interpret complex biomedical datasets, including:

- Omics data (genomics, transcriptomics, proteomics)
- Imaging data (MRI, CT, histopathology)
- Clinical records and EHRs
- In vitro and in vivo experimental measurements

Key DNN architectures to be developed and applied:

- **Multimodal Transformers**: for learning joint representations across omics, text, and image modalities.
- **Graph Neural Networks (GNNs)**: to model biological networks, signaling cascades, and drug interactions.
- **Variational Autoencoders (VAEs) and Diffusion Models**: for generating synthetic biological data and simulating patient heterogeneity.
- **Time-series Models (e.g., RNNs, Temporal Convolutional Networks)**: for longitudinal patient data and disease progression modeling.

Integration of these networks with QSP model outputs will allow:

- Augmented training of AI systems with simulated data from Virtual Patient Populations (VPpop)
- Enhanced prediction of drug responses in underrepresented populations
- Real-time adaptation to new data inputs for continuous model learning

4. Model Training, Infrastructure, and Benchmarking

Training of these AI models requires a robust infrastructure:

- **High-performance computing (HPC)** and **GPU clusters** for large-scale training and inference
- Use of federated learning techniques to ensure data privacy when using hospital/EHR data
- Integration of model training pipelines with FAIR (Findable, Accessible, Interoperable, Reusable) data repositories

Benchmarking will be conducted on:

- Standard tasks: drug response prediction, patient stratification, mechanistic hypothesis generation
- Cross-validation on real-world datasets and public biomedical AI benchmarks (e.g., BLURB, MedQA)
- Comparison with existing QSP tools and AI systems to demonstrate improvement in accuracy, usability, and interpretability

5. Human-in-the-Loop and Explainable AI

To ensure usability by clinicians and pharmacologists, a human-in-the-loop framework will be implemented:

- Interactive dashboards for exploring AI predictions
- Explanation interfaces using attention maps and counterfactuals
- Tools for expert feedback and iterative refinement of AI/QSP models

Explainability is critical for regulatory acceptance and clinical trust. Therefore, we will implement:

- Model interpretability modules (e.g., SHAP, LIME, attention visualization)
- Alignment of model outputs with biomedical ontologies and clinical decision-support rules

6. Continuous Learning and Model Evolution

Biomedical knowledge is constantly evolving. The AIQSP system will implement continuous learning mechanisms:

- Periodic re-training on newly published literature and data
- Adaptive learning strategies for concept drift and novel variants (e.g., emerging drug targets)
- Integration with active learning frameworks to prioritize the most informative data for labeling and simulation

7. Expected Impact

This extended AI framework will:

- Accelerate drug development pipelines by reducing uncertainty and improving predictions
- Lower barriers for non-specialists to use complex QSP models
- Facilitate regulatory submission by providing transparent, data-driven rationale
- Enable personalized therapy selection through multimodal patient data modeling

Project plan

1. Introduction

The AIQSP project seeks to transform biomedical research and drug development through the fusion of Quantitative Systems Pharmacology (QSP) and cutting-edge Artificial Intelligence (AI). This extension aims to elaborate on the scientific direction of the AI component, focusing on the development of next-generation generative large language models (GLLMs), deep neural networks (DNNs), and their integration with biomedical simulations and multimodal datasets.

2. Generative Large Language Models (GLLMs) for Biomedical Simulation

GLLMs, such as GPT-4 and its successors, have demonstrated remarkable abilities in natural language understanding, knowledge extraction, and reasoning. The AIQSP project proposes to develop domain-specific GLLMs tailored for biomedical applications. These models will be trained and fine-tuned on:

- Biomedical literature (e.g., PubMed, PMC)
- Clinical trial reports and regulatory filings (e.g., EMA, FDA databases)

- Drug-target interaction datasets (e.g., ChEMBL, DrugBank)
- Biological pathway and disease mechanism ontologies (e.g., Reactome, KEGG, Gene Ontology)

A specialized GLLM, termed **QSP-GPT**, will be trained using a corpus of QSP model descriptions, simulation results, and mechanistic annotations. Its primary functions will include:

- Automatic generation of hypotheses for mechanistic modeling
- Interpretation and summarization of simulation outcomes
- Suggestion of model refinements and parameter estimations
- Assistance in regulatory and clinical documentation based on model predictions

These capabilities will be enhanced with retrieval-augmented generation (RAG) mechanisms that dynamically link simulation results and structured databases during inference.

3. Deep Neural Networks for Multimodal Integration

Deep Neural Networks (DNNs) will be used to integrate and interpret complex biomedical datasets, including:

- Omics data (genomics, transcriptomics, proteomics)
- Imaging data (MRI, CT, histopathology)
- Clinical records and EHRs
- In vitro and in vivo experimental measurements

Key DNN architectures to be developed and applied:

- **Multimodal Transformers**: for learning joint representations across omics, text, and image modalities.
- **Graph Neural Networks (GNNs)**: to model biological networks, signaling cascades, and drug interactions.
- **Variational Autoencoders (VAEs) and Diffusion Models**: for generating synthetic biological data and simulating patient heterogeneity.
- **Time-series Models (e.g., RNNs, Temporal Convolutional Networks)**: for longitudinal patient data and disease progression modeling.

Integration of these networks with QSP model outputs will allow:

- Augmented training of AI systems with simulated data from Virtual Patient Populations (VPpop)
- Enhanced prediction of drug responses in underrepresented populations
- Real-time adaptation to new data inputs for continuous model learning

4. Model Training, Infrastructure, and Benchmarking

Training of these AI models requires a robust infrastructure:

- **High-performance computing (HPC)** and **GPU clusters** for large-scale training and inference
- Use of federated learning techniques to ensure data privacy when using hospital/EHR data
- Integration of model training pipelines with FAIR (Findable, Accessible, Interoperable, Reusable) data repositories

Benchmarking will be conducted on:

- Standard tasks: drug response prediction, patient stratification, mechanistic hypothesis generation
- Cross-validation on real-world datasets and public biomedical AI benchmarks (e.g., BLURB, MedQA)
- Comparison with existing QSP tools and AI systems to demonstrate improvement in accuracy, usability, and interpretability

5. Human-in-the-Loop and Explainable AI

To ensure usability by clinicians and pharmacologists, a human-in-the-loop framework will be implemented:

- Interactive dashboards for exploring AI predictions
- Explanation interfaces using attention maps and counterfactuals
- Tools for expert feedback and iterative refinement of AI/QSP models

Explainability is critical for regulatory acceptance and clinical trust. Therefore, we will implement:

- Model interpretability modules (e.g., SHAP, LIME, attention visualization)
- Alignment of model outputs with biomedical ontologies and clinical decision-support rules

6. Continuous Learning and Model Evolution

Biomedical knowledge is constantly evolving. The AIQSP system will implement continuous learning mechanisms:

- Periodic re-training on newly published literature and data
- Adaptive learning strategies for concept drift and novel variants (e.g., emerging drug targets)
- Integration with active learning frameworks to prioritize the most informative data for labeling and simulation

7. Expected Impact

This extended AI framework will:

- Accelerate drug development pipelines by reducing uncertainty and improving predictions
- Lower barriers for non-specialists to use complex QSP models
- Facilitate regulatory submission by providing transparent, data-driven rationale
- Enable personalized therapy selection through multimodal patient data modeling

9. Project Plan and Work Packages

WP1: Project Management and Coordination

- Lead: University of Edinburgh
- Objectives: Ensure timely delivery, financial reporting, partner coordination, ethics compliance
- Deliverables:
    - D1.1 Project management plan (M2)
    - D1.2 Consortium agreement and data governance structure (M3)
    - D1.3 Mid-term and final progress reports (M18, M36)
- Estimated budget: €450,000

WP2: QSP Model Development and Validation

- Lead: InsysBio
- Objectives: Build, validate and refine QSP models for breast/ovarian cancer; generate Virtual Patient Populations
- Deliverables:

    - D2.1 Validated QSP model of disease progression and drug interactions (M12)
    - D2.2 VPpop generation tool and data repository (M18)
- Estimated budget: €600,000

WP3: AI Core Development (GLLM, DNN)

- Lead: University of Edinburgh
- Objectives: Develop QSP-GPT, multimodal DNNs, and model interpretability tools
- Deliverables:
    - D3.1 Trained QSP-GPT model with biomedical tuning (M15)
    - D3.2 DNN-based multimodal integration system (M20)
    - D3.3 Model explainability dashboard (M24)
- Estimated budget: €1,200,000

WP4: Infrastructure, Integration and Continuous Learning

- Lead: TBD
- Objectives: Deploy secure HPC pipelines, integrate AI/QSP tools with FAIR databases, implement auto-updating mechanisms
- Deliverables:
    - D4.1 FAIR-compliant data platform (M9)
    - D4.2 Federated learning-enabled training pipeline (M16)
    - D4.3 Continuous learning interface for knowledge update (M30)
- Estimated budget: €800,000

WP5: Validation, Clinical Pilots and Regulatory Alignment

- Lead: Partner Clinical Institution (TBD)
- Objectives: Validate predictions in cell lines, animal models, pilot clinical trials; align outputs with EMA/FDA requirements
- Deliverables:
    - D5.1 In vitro and in vivo validation reports (M18)
    - D5.2 Pilot clinical study results (M30)

    - D5.3 Regulatory compliance documentation toolkit (M36)
- Estimated budget: €900,000

WP6: Dissemination, Exploitation, and Sustainability

- Lead: SME Partner
- Objectives: Ensure wide communication of results, develop business plan for AIQSP tools
- Deliverables:
    - D6.1 Communication and dissemination strategy (M3)
    - D6.2 Public workshops, scientific publications, open-source components (M12–36)
    - D6.3 Exploitation roadmap and sustainability strategy (M36)
- Estimated budget: €300,000

Total Estimated Budget: €4,250,000

10. Timeline Overview

- Project Start: Month 1
- Mid-Term Review: Month 18
- Final Review: Month 36

Milestones:

- M1: Project kick-off and partner onboarding (M1)
- M2: Completion of baseline QSP models (M12)
- M3: Completion of core AI systems (M20)
- M4: Completion of clinical pilots and validation (M30)
- M5: Final integration and regulatory deliverables (M36)