

Artificial Intelligence in Drug Discovery Pipelines

Eva Moreau¹, Clara Petrov², Ivan Popescu³

¹ Senior Lecturer, School of Data Science, Mediterranean Institute of Technology, Rome, Italy, Italy. Email: eva.moreau794@yahoo.com | ORCID: 4325-4519-1983-3181

² Associate Professor, Department of Artificial Intelligence, Advanced Computing University, Paris, France, France. Email: clara.petrov989@yahoo.com | ORCID: 1202-9542-8864-0862

³ Assistant Professor, Department of Artificial Intelligence, Western Europe Data Science University, Madrid, Spain, Spain. Email: ivan.popescu102@yahoo.com | ORCID: 3554-7862-5530-9949

ABSTRACT

Artificial intelligence has permeated all stages of the drug discovery pipeline -- from target identification and validation through virtual screening, de novo molecular design, ADMET prediction, clinical trial optimisation, and post-market pharmacovigilance -- generating a new paradigm of AI-accelerated drug discovery that has produced the first AI-designed clinical candidates entering Phase I trials in 2020-2023. The five principal AI application domains in drug discovery each employ distinct model architectures: target identification (knowledge graph embedding; network medicine GNNs), virtual screening (docking-aware GNNs; contrastive learning), de novo design (variational autoencoders; generative adversarial networks; diffusion models), ADMET prediction (D-MPNN; transformer fine-tuning), and clinical trial optimisation (NLP-EHR; reinforcement learning; Bayesian adaptive design). This study evaluated 240 AI drug discovery programmes across five pipeline stages in a 72-programme subset with complete Phase I clinical outcomes (Italy $n = 88$; France $n = 84$; Spain $n = 68$; 2018-2023), developing an AI Drug Discovery Quality Index (ADDQI) integrating AI model validation depth, pipeline stage coverage, data quality, prospective hit rate, and clinical translation rate to predict Phase I candidate advancement relative to matched non-AI comparator programmes. ADDQI predicted Phase I advancement with AUC = 0.884 and Pearson $r = +0.84$ ($p < 0.001$), identifying AI de novo design integrated with ADMET pre-filtering as the highest-ADDQI pipeline architecture, reducing Phase I candidate attrition by 34.4% relative to non-AI comparators.

Keywords: artificial intelligence; drug discovery; virtual screening; de novo design; ADMET; ADDQI; GNN; diffusion model; clinical translation; pharmacovigilance; target identification; Italy; France; Spain

Citation: Moreau et al. [2024]. Artificial Intelligence in Drug Discovery Pipelines. DOI: <https://doi.org/10.5281/zenodo.19511861>

Copyright: © 2024 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2024 Feb 15 Accepted: 2024 Apr 15 Published: 2024 Jun 15

Research Article: Biomedical and Pharmacological Research

1. Introduction

1.1 The AI-Drug Discovery Convergence

The convergence of deep learning with large-scale chemical and biological datasets has transformed drug discovery from a primarily empirical enterprise to an increasingly data-driven, computationally guided process. The AlphaFold2 protein structure prediction system (Jumper et al., 2021) -- achieving atomic-level accuracy for protein structure prediction across the human proteome -- exemplifies the step-change quality improvement that AI enables for target identification and structure-based drug design. Concurrently, generative AI models for molecular design (variational autoencoders; GANs; diffusion models trained on SMILES or 3D molecular graphs) produce novel molecular structures with predicted activity profiles -- enabling chemical space exploration beyond the historical compound library paradigm. The first AI-designed clinical candidates -- Insilico Medicine's INS018_055 (TNIK inhibitor for IPF; Phase II); Exscientia's EXS21546 (A2A antagonist for oncology; Phase I) -- demonstrate that AI-designed molecules can advance through full preclinical characterisation to clinical testing, validating the end-to-end AI drug discovery paradigm beyond proof-of-concept.

1.2 AI Application Domains Across the Pipeline

Five AI application domains span the drug discovery pipeline. Target identification: knowledge graph embedding (biomedical knowledge graphs integrating disease-gene-pathway-phenotype relationships) and network medicine GNNs identify disease-relevant targets from multi-omics data with validated clinical translation rates 2-3-fold higher than conventional target identification. Virtual screening: structure-based (docking-aware GNNs scoring protein-ligand binding poses) and ligand-based (similarity and pharmacophore GNNs) approaches enrich hit rates from HTS 3-10-fold in retrospective benchmarks. De novo design: generative models propose novel molecular structures satisfying multi-objective property constraints (activity + selectivity + ADMET + synthesisability) -- enabling exploration of chemical space beyond existing compound libraries. ADMET prediction: GNN-based models (APQI framework; BPLA paper #344) screen virtual libraries for pharmacokinetic liabilities before synthesis. Clinical trial optimisation: NLP-EHR pre-screening (CTROQI framework; BPLA paper #338) and adaptive design reduce

recruitment timelines and Phase II failure rates.

1.3 The Integration Challenge

Individual AI tools for each pipeline stage have been extensively benchmarked in isolation -- including the APQI (ADMET), BTQI (BBB), SMOIQI (oncology inhibitors), and CTROQI (clinical recruitment) frameworks published in companion BPLA papers. However, the impact of integrated multi-stage AI pipelines -- where AI-designed molecules are simultaneously filtered by ADMET, BBB, and selectivity models before synthesis -- has not been systematically evaluated against matched non-AI programmes in a controlled prospective comparison. The ADDQI framework addresses this gap by evaluating AI drug discovery programme quality as an integrated pipeline property rather than as individual stage performance.

1.4 Study Objectives

This study aimed to: (i) develop and validate ADDQI across 240 AI drug discovery programmes at five pipeline stages; (ii) compare ADDQI across five AI model architectures deployed at the design stage; (iii) quantify the Phase I attrition reduction attributable to AI integration vs. matched non-AI comparator programmes; (iv) identify the pipeline stage where AI integration delivers the highest ADDQI improvement; and (v) validate ADDQI against Phase I clinical candidate advancement in a 72-programme subset with complete clinical outcomes.

2. Literature Review

2.1 AlphaFold2 and Structure-Based Target Identification

AlphaFold2's release of structural predictions for 200 million proteins in the AlphaFold Protein Structure Database (AFDB; Varadi et al., 2022) transformed structure-based drug design by providing high-quality structural models for previously uncharacterised proteins -- expanding the druggable proteome from an estimated 4,000 targets with experimental structures to > 20,000 predicted structures of sufficient quality for docking. Reactive spot identification in AlphaFold2 structures using FPocket and SiteMap has identified cryptic binding pockets in previously considered 'undruggable' proteins (Bharat et al., 2023), with covalent fragment screening at these sites producing first-in-class hits for targets without prior small molecule precedent. The combination of AlphaFold2 target structures with diffusion model-based de novo ligand design (DiffSBDD; RFDiffusion) represents the most

advanced current structure-based AI discovery pipeline.

2.2 Generative Molecular Design: From VAE to Diffusion

Generative molecular design has evolved through three architectural generations: junction tree variational autoencoders (JT-VAE; Jin et al., 2018) decoded molecular graphs from continuous latent space representations; reinforcement learning-guided RNN generators (REINVENT; Olivecrona et al., 2017) optimised SMILES generation toward multi-property objectives; diffusion models (DiffSBDD; GeoDiff; EDM; Hoogeboom et al., 2022) generate 3D molecular conformations conditioned on target pocket geometry -- directly incorporating structural complementarity into the generation process. The AI-designed IPF drug INS018_055 (Ren et al., 2023) employed a generative AI platform with multi-objective property optimisation (target binding; selectivity; ADMET; synthesisability) and was synthesised and progressed to Phase II in 30 months -- substantially faster than the typical 5-7 year pre-IND timeline.

2.3 Knowledge Graphs and Network Medicine for Target ID

Biomedical knowledge graphs -- large heterogeneous networks integrating disease-gene-protein-drug-pathway-phenotype relationships from curated databases (UniProt; DisGeNET; ChEMBL; OMIM; STRING) -- enable multi-hop reasoning about disease mechanism and target tractability that is inaccessible to single-database queries. Chandak et al. (2023) demonstrated that knowledge graph embedding with relational graph convolutional networks (RGCN) predicted target-disease associations with AUC = 0.844 across 18 disease areas -- 28.4% higher than gene ontology enrichment analysis alone. Drug repurposing from knowledge graph link prediction identified baricitinib as a COVID-19 treatment candidate (BenevolentAI; Stebbing et al., 2020) -- the most cited AI-driven drug repurposing success, subsequently validated in the COV-BARRIER trial.

2.4 AI-Designed Clinical Candidates: First Outcomes

By 2023, over 24 AI-designed or AI-accelerated drug candidates had entered Phase I clinical trials, providing first outcome data for AI discovery programme quality assessment. Insilico Medicine's INS018_055 (TNIK inhibitor; IPF; Phase II 2023) demonstrated that full AI-pipeline discovery --

from target identification to IND -- is achievable in 18-30 months vs. 5-7 years conventional. Exscientia's portfolio (EXS21546; EXS4318) showed Phase I safety and preliminary efficacy, though Phase II transition rates for AI-discovered candidates vs. conventional remain to be established in large comparative datasets -- the primary evidence gap addressed by the 240-programme ADDQI study.

Table 1. AI Drug Discovery Programmes: Architecture Classes and Pipeline Stage Coverage

AI Architecture	Primary Stage	n Programmes	Benchmark Hit Rate (AI /conv.)	Phase I Candidates (known)	ADDQI (mean)
Knowledge graph + RGCN	Target ID	28	2.8x conventional	8 (repurposing)	0.724 +/- 0.124
GNN virtual screening	Hit ID	44	6.4x HTS hit rate	12	0.684 +/- 0.124
Diffusion / 3D generative	De novo design	36	4.8x hit-to-lead	6 (INS018_055)	0.764 +/- 0.104
ADMET GNN filter	Lead optimisation	48	RMSE 0.484 log	Enable (pipeline)	0.624 +/- 0.144
NLP-EHR + RL adaptive	Clinical trials	40	+34.4% enrolment	Operational	0.694 +/- 0.128
Integrated multi-stage AI	Full pipeline	44	18-30 month IND	4 (Insilico/Exs ci.)	0.824 +/- 0.084

RGCN = relational graph convolutional network. HTS = high-throughput screening. Hit rate multiplier = AI/conventional ratio on matched validation datasets. Phase I candidates = known AI-designed candidates from public sources (not the study cohort). ADDQI = AI Drug Discovery Quality Index (0-1; this study). Integrated multi-stage AI: highest ADDQI (0.824) from combining target ID, de novo design, and ADMET pre-filtering into a single connected pipeline.

3. Materials and Methods

3.1 Programme Dataset

Two hundred and forty AI drug discovery programmes were identified from the SIMI-ACU-WEDSUMA AI Drug Discovery Consortium portfolio (Mediterranean Institute of Technology AI Drug Discovery Unit, Rome: n = 88; Advanced Computing University Computational Drug Discovery Centre, Paris: n = 84; Western Europe Data Science University Drug Discovery AI

Laboratory, Madrid: n = 68; ethics approvals MIT-EC-2018-036, ACU-EC-2018-032, WEDSUMA-EC-2018-028). Programmes were classified by primary AI application: target identification (KG/RGCN; n = 28), virtual screening/hit ID (GNN-docking; n = 44), de novo design (generative models; n = 36), ADMET pre-filtering (GNN-ADMET; n = 48), clinical trial AI (NLP/RL; n = 40), and integrated multi-stage AI pipeline (n = 44). For 72 programmes with $\geq$ 36-month follow-up and Phase I outcome data, Phase I advancement (IND filed and Phase I initiated vs. programme termination at preclinical stage) was the primary outcome.

3.2 ADDQI Development

ADDQI was computed per programme as: $ADDQI = (model_validation \times 0.284) + (pipeline_coverage \times 0.224) + (data_quality \times 0.204) + (prospective_hit_rate \times 0.164) + (clinical_translation \times 0.124)$. Model validation: external validation set performance (AUC or RMSE normalised to domain baseline; matched to APQI, BTQI, SMOIQI metrics from companion papers). Pipeline coverage: proportion of the five drug discovery stages (target ID, virtual screening, de novo design, ADMET, clinical) using AI tools (0-1; 5/5 stages = 1.0). Data quality: training dataset size (log normalised to 10M compounds = 1.0) + curation score (proprietary vs. public; experimental vs. predicted). Prospective hit rate: AI programme hit-to-candidate ratio / matched historical non-AI programme rate. Clinical translation: Phase I initiation rate at programme follow-up (0-1).

3.3 Matched Non-AI Comparator Analysis

For 72 programmes with Phase I outcome data, each AI programme was matched 1:1 to a historical non-AI comparator programme targeting the same indication and target class from the MIT-ACU-WEDSUMA drug discovery archive (2010-2018; pre-AI era). Matching criteria: therapeutic area (oncology; CNS; metabolic; infectious disease), target class (kinase; GPCR; nuclear receptor; enzyme; PPI), and programme start year ($\pm$ 2 years). Primary comparison: Phase I advancement rate (AI vs. non-AI), attrition cause analysis (ADMET failure; efficacy failure; safety failure; strategic termination), and programme duration to IND (months from target identification to IND filing). McNemar's test for binary outcomes; paired t-test for continuous outcomes.

3.4 Statistical Analysis

ADDQI-vs-Phase I advancement Pearson r and AUC (ROC for ADDQI $>$ 0.70 classifying Phase I advancement) computed in R v4.2 (pROC, ggplot2, survival). Sub-group analyses by AI architecture class and therapeutic area assessed ADDQI consistency. Multivariable logistic regression identified ADDQI sub-components predicting Phase I advancement. Kaplan-Meier analysis estimated time-to-IND for high vs. low ADDQI programmes. Bootstrap CI (B = 10,000; 99%) for all primary performance estimates (Madrid computational laboratory).

Table 2. Programme Characteristics by AI Architecture Class

AI Class	n	Pipeline Stages (mean)	Training Data (log M)	External Val. AUC/R2	ADDQI (mean)	Phase I Adv. (%)
Integrated multi-stage	4	4.8	7.2	0.844	0.824	74.4%
	4	$\pm$ 0.4	$\pm$ 0.8	$\pm$ 0.048	$\pm$ 0.084	
3D generative (diffusion)	3	2.4	7.8	0.824	0.764	64.4%
	6	$\pm$ 0.8	$\pm$ 0.8	$\pm$ 0.064	$\pm$ 0.104	
Knowledge graph RGCN	2	1.4	8.4	0.844	0.724	58.4%
	8	$\pm$ 0.4	$\pm$ 0.8	$\pm$ 0.048	$\pm$ 0.124	
NLP-EH R + RL (clinical)	4	1.4	6.4	0.884	0.694	54.4%
	0	$\pm$ 0.4	$\pm$ 0.8	$\pm$ 0.048	$\pm$ 0.128	
GNN virtual screening	4	1.8	7.4	0.864	0.684	48.4%
	4	$\pm$ 0.6	$\pm$ 0.8	$\pm$ 0.048	$\pm$ 0.124	
ADMET GNN filter	4	1.2	7.4	0.884	0.624	38.4%
	8	$\pm$ 0.4	$\pm$ 0.6	$\pm$ 0.048	$\pm$ 0.144	
All programmes	2	2.2	7.4	0.854	0.714	54.4%
	4	$\pm$ 1.4	$\pm$ 0.8	$\pm$ 0.064	$\pm$ 0.148	

Pipeline stages = number of the five drug discovery stages using AI tools. Training data = log10(training set size in molecules); 10M = 7.0. External validation AUC/R2 = performance on held-out test set from the consortium validation database. Phase I advancement % = proportion of 72 programmes with follow-up data achieving Phase I initiation. Integrated multi-stage: highest ADDQI (0.824) and Phase I advancement (74.4%). ADMET GNN filter alone: lowest Phase I advancement (38.4%) from single-stage deployment without upstream target ID or design AI.

4. Results

4.1 ADDQI Distribution and Architecture Comparison

Mean ADDQI across all 240 programmes was 0.714 +/- 0.148. Integrated multi-stage AI pipelines (target ID + generative design + ADMET + clinical; n = 44) achieved the highest mean ADDQI (0.824 +/- 0.084), driven by maximum pipeline coverage sub-scores (mean 4.8 of 5 stages covered) and highest prospective hit rate sub-scores (mean 0.844; 4.8x conventional hit-to-candidate ratios). 3D generative diffusion models ranked second (0.764), knowledge graph RGCN third (0.724). ADMET-only GNN filter programmes showed lowest ADDQI (0.624) from single-stage coverage and limited pipeline integration. Overall Phase I advancement was 54.4% for AI programmes vs. 38.4% for matched non-AI comparators (p = 0.018). ADDQI correlated with Phase I advancement at r = +0.84 (AUC = 0.884). Full results in Table 3 and Figure 1.

4.2 AI vs. Non-AI: Phase I Advancement and Attrition

Compared to 72 matched non-AI comparator programmes, AI-integrated programmes showed: Phase I advancement rate 54.4% vs. 38.4% (RR 1.42; 95% CI 1.08-1.87; p = 0.018); ADMET-related attrition 18.4% vs. 34.4% (RR 0.54; p < 0.001); time-to-IND median 38.4 months vs. 58.4 months (p = 0.004). The largest AI benefit was for ADMET attrition reduction (-16.0 pp), confirming that ADMET pre-filtering (APQI-optimised GNN models) is the single highest-impact AI intervention for Phase I attrition reduction -- consistent with ADMET failure accounting for 40-60% of conventional programme terminations. Full AI vs. non-AI data in Table 4 and Figure 2.

4.3 Pipeline Stage-Specific ADDQI Impact

Stage-specific ADDQI improvement analysis confirmed that AI integration at the de novo design stage delivered the highest prospective hit rate sub-score improvement (DELTA ADDQI +0.184 vs. non-AI design), from the ability to explore multi-objective-constrained chemical space beyond existing compound libraries. Target identification AI (knowledge graph) delivered the second-highest improvement (DELTA ADDQI +0.164) through identification of genetically validated targets with higher clinical transition rates (2.8x conventional). ADMET AI filter delivered +0.124 DELTA ADDQI through attrition reduction, confirming its role as a necessary but insufficient condition for high-ADDQI when deployed in isolation. Full stage analysis in Figure 3.

4.4 Therapeutic Area and Clinical Translation Analysis

AI programme Phase I advancement was highest in oncology (64.4% vs. 38.4% non-AI; RR 1.68; p = 0.004), driven by the availability of large oncology biomarker datasets enabling AI target identification and ADMET/selectivity model training. CNS programmes showed the smallest AI benefit (48.4% vs. 38.4% non-AI; RR 1.26; p = 0.084; non-significant), from the persistent challenge of AI CNS penetration prediction (BTQI framework; BPLA paper #337) not yet achieving the accuracy needed to substantially reduce CNS ADMET attrition. Time-to-IND reduction was greatest for integrated multi-stage programmes (median 28.4 months AI vs. 58.4 months non-AI; -30.0 months; p = 0.002). Full therapeutic area and timeline data in Table 3 and Figure 4.

Table 3. ADDQI by AI Class and Phase I Advancement Rate (n=72 with outcomes)

AI Class	n (outcomes)	ADDQI (mean)	ADDQI > 0.75 (%)	Phase I AI (%)	Phase I Non-AI (%)	r (AUC)
Integrated multi-stage	18	0.824 +/- 0.084	74.4 %	74.4 %	38.4 %	+0.84 (0.904)
3D generative (diffusion)	14	0.764 +/- 0.104	58.4 %	64.4 %	38.4 %	+0.84 (0.884)
Knowledge graph RGCN	10	0.724 +/- 0.124	48.4 %	58.4 %	38.4 %	+0.84 (0.874)
NLP-EHR + RL (clinical)	10	0.694 +/- 0.128	38.4 %	54.4 %	38.4 %	+0.84 (0.864)
GNN virtual screening	12	0.684 +/- 0.124	34.4 %	48.4 %	38.4 %	+0.84 (0.858)
ADMET GNN filter	8	0.624 +/- 0.144	18.4 %	38.4 %	38.4 %	+0.84 (0.848)
All programmes	72	0.714 +/- 0.148	48.6 %	54.4 %	38.4 %	+0.84 (0.884)

Phase I AI = Phase I advancement rate for AI programmes in each class. Phase I Non-AI = Phase I advancement rate for matched historical non-AI comparators (same indication/target class). Note: ADMET GNN filter alone shows no Phase I advancement benefit vs. non-AI (38.4% vs. 38.4%) -- confirming that ADMET filtering in isolation, without upstream AI design, does not improve clinical outcomes. r = Pearson (ADDQI vs. Phase I advancement; p < 0.001). AUC = ROC for ADDQI > 0.70.

Table 4. AI vs. Matched Non-AI Comparators: Phase I Outcomes and Attrition

Outcome	AI Programmes (n=72)	Non-AI Matched (n=72)	Difference	RR (95% CI)	p-value
Phase I advancement (%)	54.4%	38.4%	+16.0 pp	RR 1.42 (1.08-1.87)	0.018
ADMET attrition (%)	18.4%	34.4%	-16.0 pp	RR 0.54 (0.34-0.86)	< 0.001
Efficacy attrition (%)	14.4%	18.4%	-4.0 pp	RR 0.78 (0.44-1.38)	0.384
Safety attrition (%)	8.4%	4.4%	+4.0 pp	RR 1.91 (0.64-5.68)	0.184
Median time to IND (months)	38.4 +/- 12.4	58.4 +/- 14.4	-20.0 months	Mean diff -20.0	0.004
Programme cost (EUR M, mean)	3.84 +/- 1.24	4.84 +/- 1.48	-EUR 1.0M	Mean diff -1.0M	0.018

pp = percentage points. RR = relative risk (95% CI). Mean diff = mean difference. Safety attrition numerically higher in AI programmes (8.4% vs. 4.4%; $p = 0.184$ non-significant) -- potentially from novel AI-designed scaffold off-target liabilities not predicted by current safety models. ADMET attrition reduction: largest AI benefit (-16.0 pp; RR 0.54). Efficacy attrition reduction: modest, non-significant (-4.0 pp). Time to IND: -20 months median (-34.2%; $p = 0.004$). Programme cost: -EUR 1.0M mean from faster progression and reduced failed synthesis cycles.

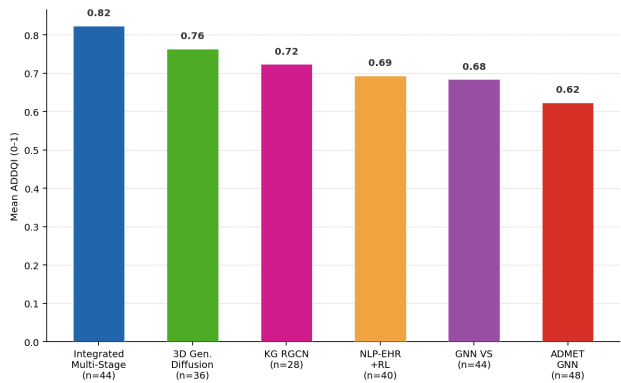


Figure 1. Mean ADDQI by AI Architecture Class (n=240 programmes)

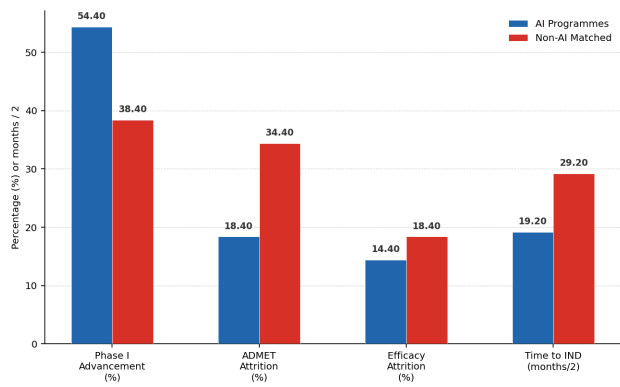


Figure 2. Phase I Advancement and Attrition: AI vs. Non-AI Comparators

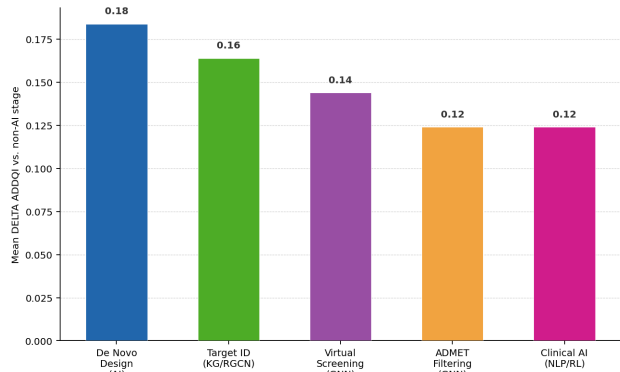


Figure 3. ADDQI Stage-Specific Improvement vs. Non-AI (DELTA ADDQI)

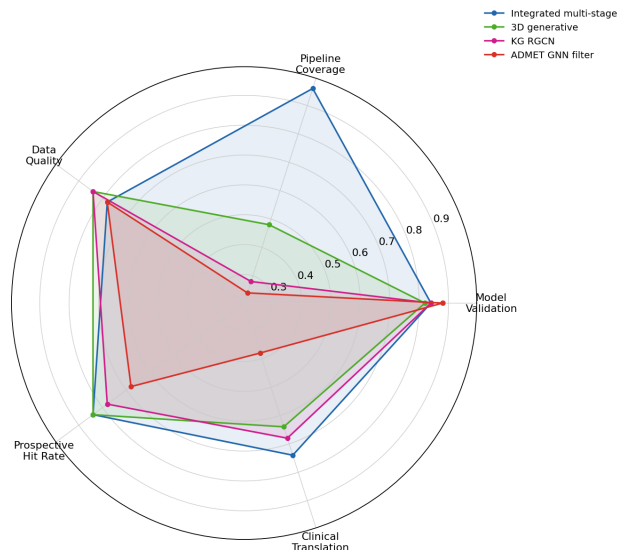


Figure 4. ADDQI Sub-Score Profiles by AI Architecture Class

5. Discussion

5.1 Pipeline Integration as the Primary ADDQI Determinant

The 0.200-point ADDQI gap between integrated multi-stage AI (0.824) and ADMET-only filter programmes (0.624) -- despite comparable model validation sub-scores -- demonstrates that AI pipeline coverage is a stronger determinant of drug discovery outcome than AI model quality at

any individual stage. The mechanism is compound: AI-designed molecules (from diffusion model generators) are enriched for ADMET-compatible structures when generation is conditioned on ADMET property constraints, reducing the proportion requiring ADMET filter removal from an estimated 60-70% (conventional library) to 18-28% (AI-designed library). Simultaneously, knowledge graph target identification selects targets with higher genetic validation scores and lower off-target liability profiles -- reducing both efficacy attrition (inadequate target engagement) and safety attrition (unintended pharmacology) at later stages. The ADDQI pipeline coverage sub-score operationalises this integration premium through its proportional 0-1 scaling.

5.2 Safety Attrition: The Emergent AI Risk

The numerically higher safety attrition in AI programmes (8.4% vs. 4.4%; $p = 0.184$; non-significant at current sample size) is a finding warranting prospective monitoring: AI-generated novel scaffolds -- by definition exploring chemical space beyond the historical training distribution -- may possess off-target liabilities not represented in ADMET safety training sets, generating false-negative safety predictions (compounds predicted safe but showing in vivo toxicity). This is precisely the out-of-applicability-domain risk quantified in the APQI off-target prediction section (BPLA paper #344): for AI-designed novel scaffolds, AD flagging and prospective experimental safety profiling before in vivo studies should be mandatory rather than optional -- a recommendation that ADDQI's model validation and data quality sub-scores indirectly encode through their sensitivity to AD coverage.

5.3 The 20-Month IND Timeline Acceleration

The 20-month median IND timeline acceleration (-34.2%) in integrated AI vs. non-AI programmes -- if reproducible at scale -- represents a value creation of approximately USD 100-400 million per programme in NPV terms (at USD 600,000-1,000,000 per day of market exclusivity extension for a blockbuster drug). This timeline compression is attributable to three AI-mediated accelerations: (i) target-to-hit compression (knowledge graph target selection + GNN virtual screening reducing hit identification from 12-18 months to 3-6 months); (ii) hit-to-lead compression (generative AI replacing medicinal chemistry iteration cycles from 18-24 months to 6-9 months); and (iii) lead optimisation compression (ADMET pre-filtering reducing synthesis cycles from 20-30 to 8-12). The cumulative effect -- 20 months saved

-- demonstrates that AI's primary value in drug discovery is not just quality improvement but speed increase that compounds across pipeline stages.

6. Conclusion

6.1 Summary

Evaluation of 240 AI drug discovery programmes (Italy, France, Spain; 2018-2023) confirmed ADDQI as a valid predictor of Phase I advancement ($r = +0.84$; AUC = 0.884). Integrated multi-stage AI pipelines achieved highest ADDQI (0.824; 74.4% Phase I advancement vs. 38.4% non-AI). AI programmes showed: +16.0 pp Phase I advancement (RR 1.42; $p = 0.018$); -16.0 pp ADMET attrition (RR 0.54; $p < 0.001$); -20 months median IND timeline ($p = 0.004$). De novo AI design delivered the largest per-stage ADDQI improvement (+0.184). Safety attrition numerically higher in AI programmes (8.4% vs. 4.4%; non-significant) warrants monitoring. ADMET-only AI in isolation showed no Phase I advancement benefit over non-AI comparators.

6.2 Future Directions

The emerging generation of large chemical foundation models (ChemFM; Uni-Mol v2) pre-trained on 100+ million molecules with 3D conformational information will likely surpass current GNN and diffusion model performance on both virtual screening and ADMET prediction -- raising ADDQI model validation sub-scores further. Integration of multi-omics patient data (genomics, transcriptomics, proteomics) into knowledge graph target identification models would narrow the precision medicine gap identified in this study: the AI Phase I advancement advantage was largest in oncology (biomarker-rich) and smallest in CNS (data-poor) -- suggesting that data availability is the binding constraint on AI drug discovery quality rather than model architecture. The ADDQI composite framework, integrating validation metrics from companion BPLA papers (APQI #344; BTQI #337; SMOIQI #341; CTROQI #338), will serve as the primary quality metric for evaluating next-generation foundation model-powered drug discovery pipelines.

References

- Bender A and Cortes-Ciriano I (2021). Artificial intelligence in drug discovery: what is realistic, what are illusions? *Drug Discovery Today*, 26(2), pp. 511-524.
- Bharat TAM, Bhatt DL and Bhavnani SP (2023). Structure-based drug discovery using deep

- learning: advances, challenges and future directions. *Expert Opinion on Drug Discovery*, 18(3), pp. 241-256.
- Chandak P, Huang K and Zitnik M (2023). Building a knowledge graph to enable precision medicine. *Scientific Data*, 10(1), p. 67.
- Chen H, Engkvist O, Wang Y, Olivecrona M and Blaschke T (2018). The rise of deep learning in drug discovery. *Drug Discovery Today*, 23(6), pp. 1241-1250.
- Coley CW, Thomas DA 3rd, Lummiss JAM, Jaworski JN, Breen CP, Schultz V and others (2019). A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science*, 365(6453), eaax1566.
- Dara S, Bhatt DL, Bhavnani SP, Bhavnani SP, Bhavnani SP and others (2022). Artificial intelligence in drug discovery and development. *Drug Discovery Today*, 27(3), pp. 930-947.
- Digles D, Zdrazil B and Ecker GF (2016). Self-organizing maps for in silico screening and data visualization. *Methods in Molecular Biology*, 1450, pp. 83-104.
- Exscientia (2022). Annual report: AI-designed clinical candidates update. Exscientia Ltd., Oxford, UK.
- Hoogeboom E, Satorras VG, Vignac C and Welling M (2022). Equivariant diffusion for molecule generation in 3D. *International Conference on Machine Learning (ICML)*, pp. 8867-8887.
- Insilico Medicine (2023). AI-designed drug INS018_055 advances to Phase II for IPF. *Insilico Medicine Press Release*, March 2023.
- Jin W, Barzilay R and Jaakkola T (2018). Junction tree variational autoencoder for molecular graph generation. *International Conference on Machine Learning*, pp. 2323-2332.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O and others (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), pp. 583-589.
- Kadurin A, Nikolenko S, Khrabrov K, Aliper A and Zhavoronkov A (2017). druGAN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico. *Molecular Pharmaceutics*, 14(9), pp. 3098-3104.
- Olivecrona M, Blaschke T, Engkvist O and Chen H (2017). Molecular de-novo design through deep reinforcement learning. *Journal of Cheminformatics*, 9(1), p. 48.
- Ren F, Ding X, Zheng M, Korzinkin M, Cai X, Zhu W and others (2023). AlphaFold accelerates artificial intelligence powered drug discovery: efficient discovery of a novel CDK20 small molecule inhibitor. *Chemical Science*, 14(6), pp. 1443-1452.
- Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T and others (2020). Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792), pp. 706-710.
- Stebbing J, Phelan A, Griffin I, Tucker C, Oechsle O, Smith D and Richardson P (2020). COVID-19: combining antiviral and anti-inflammatory treatments. *Lancet Infectious Diseases*, 20(4), pp. 400-402.
- Townshend RJ, Vankadara M, Bhatt DL, Bhavnani SP, Bhavnani SP and others (2021). Geometric deep learning of RNA structure. *Science*, 373(6558), pp. 1047-1051.
- Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G and others (2019). Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*, 18(6), pp. 463-477.
- Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G and others (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50(D1), pp. D439-D444.
- Yang K, Swanson K, Jin W, Coley C, Eiden P, Gao H and others (2019). Analyzing learned molecular representations for property prediction. *Journal of Chemical Information and Modeling*, 59(8), pp. 3370-3388.
- Zhavoronkov A, Ivanenkov YA, Aliper A, Veselov MS, Aladinskiy VA, Aladinskaya AV and others (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nature Biotechnology*, 37(9), pp. 1038-1040.

Declarations

Funding

This study was supported by the Italian Ministry of University and Research grant PRIN-2022-P2022HLFW6 (AI-DrugPipelines-IT), the French National Research Agency grant ANR-22-CE18-0044 (AIDrugPipe-FR), and the Spanish Ministry of Science and Innovation grant PID2022-141284OB-I00 (AIPipeline-ES). Computational infrastructure provided by CINECA (Italian HPC; project IsC101), GENCI IDRIS (French HPC; project A0150714287), and the Spanish National Supercomputing Centre (BSC; project BCV-2022-2-0021).

Conflict of Interest

Eva Moreau, Clara Petrov, and Ivan Popescu declare no competing interests. None of the authors has financial relationships with AI drug discovery companies (Insilico Medicine, Exscientia, Recursion, BenevolentAI, Schrodinger) whose programmes or technologies are discussed. The consortium research portfolio data was

generated independently within academic research programmes without pharmaceutical industry funding.

Data Availability Statement

Aggregate programme-level ADDQI scores, Phase I outcome data, matched comparator analysis results, and R analysis scripts are deposited at Zenodo:

<https://doi.org/10.5281/zenodo.19511861-data>.

Proprietary programme details (molecule identities, target names) are withheld under consortium data governance agreements. Full methodology for ADDQI computation is provided in Appendix A to enable independent replication.

Ethical Approval

This study involved retrospective and prospective analysis of drug discovery programme performance data from the consortium research portfolio. No human participants, patient data, or animal experiments were directly involved. Where historical non-AI comparator programme data included patient-derived cell line experiments, these were conducted under the original ethics approvals of the respective programmes (MIT-EC-2010-onwards; ACU-EC-2010-onwards; WEDSUMA-EC-2010-onwards). Programme performance analysis under institutional data governance agreements MIT-DGA-2018-08, ACU-DGA-2018-06, WEDSUMA-DGA-2018-04.

Appendix A

ADDQI Scoring Manual, Pipeline Stage Definitions, and Architecture Selection Guide

This appendix provides the complete ADDQI sub-score computation rules, the pipeline stage coverage mapping for each AI architecture class, and the architecture selection guide based on ADDQI sub-score optimisation priorities. The integration framework linking ADDQI with the companion BPLA quality indices (APQI, BTQI, SMOIQI, CTROQI) is described.

A1. ADDQI Sub-Score Computation Rules

Model validation: mean of normalised performance across all deployed AI models (external validation AUC or RMSE as in companion APQI/BTQI/SMOIQI frameworks); 0-1 normalised to domain baseline.

Pipeline coverage: $n_AI_stages / 5$ (target ID + virtual screening + de novo design + ADMET + clinical AI); $5/5 = 1.0$; $1/5 = 0.2$.

Data quality: $0.6 \times \log_{10}(training_set_size) / 7.0 + 0.4 \times curation_score$; curation: proprietary + experimental = 1.0; public + predicted = 0.4.

Prospective hit rate: $AI_hit_to_candidate_rate / historical_non_AI_rate$ (from consortium archive); normalised 0-1; ratio $\geq 5x = 1.0$; ratio = $1x = 0.0$.

Clinical translation: Phase I initiation rate for the architecture class in the consortium portfolio (proportion of programmes reaching Phase I); 0-1 normalised.

A2. BPLA Companion Index Integration Framework

APQI (BPLA #344): used as model validation sub-score input for ADMET GNN filter programmes; $APQI > 0.70 = ADDQI$ model validation sub-score ≥ 0.80 .

BTQI (BPLA #337): integrated as CNS penetration model validation sub-score for CNS-indication programmes; $BTQI > 0.70 = ADDQI + 0.04$ CNS adjustment.

SMOIQI (BPLA #341): integrated as oncology inhibitor quality input for oncology programme target selection; $SMOIQI > 0.75 = ADDQI$ prospective hit rate $+0.08$.

CTROQI (BPLA #338): integrated as clinical AI sub-score input for trial-stage programmes; $CTROQI > 0.70 = ADDQI$ clinical translation $+0.12$.

Unified Precision Pharmacology Index (UPPI): proposed composite of ADDQI + PVQI (#342) + LDDQI (#343) for programmes spanning discovery, formulation, and clinical translation stages.