

ToxTwin — Synthèse technique consolidée

Version V2.3 · Avril 2026 · Twingital Institute / Qualees

1. Vue d'ensemble

ToxTwin est une plateforme de scoring toxicologique prédictif pré-Phase 1, fondée sur un pipeline GNN (Graph Neural Network) end-to-end. Elle couvre 14 endpoints toxicologiques (12 Tox21 + Ames mutagénicité + hERG inhibition), complétés par une couche d'interprétation pharmacologique par LLM local. L'ensemble s'exécute sur une infrastructure souveraine sans dépendance cloud.

La plateforme a traversé un cycle de développement en six versions majeures (V1.0→V2.3), marqué par la découverte et la correction rigoureuse de biais méthodologiques. La version V2.3 constitue la première version dont les métriques sont validées sur un protocole 5-fold scaffold CV strict.

Métriques clés V2.3 :

Indicateur	Valeur	Protocole
Tox21 Mean AUC	0.867 ± 0.043	5-fold scaffold CV
Ames AUC	0.843 ± 0.029	5-fold scaffold CV
hERG AUC	0.785 ± 0.053	5-fold scaffold CV
Cibles atteintes (≥ seuil)	12/14	5-fold scaffold CV
Calibration ECE (14 ep.)	< 0.05	Régression isotonique OOF
Endpoints	14	12 Tox21 + Ames + hERG

2. Trajectoire V1.0 → V2.3 — Biais, corrections, gains

2.1 Chronologie

Version	Période	Chantier	Résultat clé
V1.0	Jan 2026	Architecture GNN initiale, pré-entraînement, fine-tuning Tox21	AUC 0.857 annoncée — non reproductible
V1.1	Fév 2026	Calibration, domaine applicatif	

Version	Période	Chantier	Résultat clé
			ECE 0.043 PASS — AD sur embeddings aléatoires
V1.2	Fév-Mar 2026	Extension hERG + Ames, Focal Loss	AUC Ames 0.968 — circulaire
V1.3	Mar-Avr 2026	Audit complet, correction d'architecture, labels Hansen	AUC réelle : 0.594 ± 0.056
V2.0	Avr 2026	Nouveau corpus, pré-entraînement, fine-tuning multi-tâche	Tox21 0.722, Ames 0.769, hERG 0.691
V2.1	Avr 2026	Optimisation d'entraînement	Gain marginal
V2.2	Avr 2026	Architecture ensemble multi-représentation	Gains sélectifs par endpoint
V2.3	Avr 2026	Mécanisme de sélection par endpoint + calibration V2.3	Tox21 0.867, 12/14 cibles atteintes

2.2 Biais découverts et corrigés

Bug fondamental d'architecture (V1.0-V1.2). Incompatibilité silencieuse entre le checkpoint pré-entraîné et l'architecture d'inférence. Le chargement avec `strict=False` masquait le problème : les poids étaient ignorés, produisant des embeddings sans signification. Conséquence : tous les scores V1.0-V1.2 étaient calculés sur des poids aléatoires.

Fuite de données (V1.0-V1.1). Contamination partielle du val set par migration de composés vers le train. Split aléatoire initial (non scaffold) : surestimation estimée 3-8 % (Sheridan 2013). AUC 0.857 mesurée during-training sur val partiellement contaminé.

Circularité Ames (V1.2). AUC Ames 0.968 = artefact structurel. Labels SMARTS encodent des alertes de sous-structures que le GNN apprend à reproduire (règle de génération, pas phénotype biologique). AUC réelle sur données expérimentales Hansen : 0.560 ± 0.036 .

AD dégénérée (V1.1-V1.2). Composantes du domaine applicatif calculées dans un espace de dimensionnalité inadaptée et avec des seuils auto-référencés.

2.3 Mesures correctives

Phase 0 — Fondations. Scaffold split Bemis-Murcko strict sur l'ensemble du corpus. Vérification InChIKey : train n val n test = \emptyset . Holdout test gelé (usage unique). Baseline 5-fold CV établie.

Phase 1 — Corpus et encodeur (V2.0). Remplacement du corpus de pré-entraînement par un corpus drug-like de référence. Re-pré-entraînement et fine-tuning multi-tâche 14 endpoints.

Phase 2 — Architecture ensemble (V2.2). Fusion de représentations moléculaires complémentaires. L'architecture de fusion, les dimensionnalités et les mécanismes de projection constituent une propriété intellectuelle Twingital Institute.

Phase 3 — Sélection par endpoint (V2.3). Le mécanisme de routage, la logique de sélection et la table d'affectation constituent une propriété intellectuelle Twingital Institute.

3. Architecture V2.3

3.1 Composants

L'architecture V2.3 combine plusieurs représentations moléculaires (topologique, attentionnelle, sous-structurale) via un mécanisme de fusion et de sélection par endpoint.

Les spécifications détaillées des composants — dimensionnalités, statuts d'entraînement, nombre de paramètres — constituent une propriété intellectuelle Twingital Institute.

3.2 Sélection par endpoint

La table de routage, les critères de sélection et la répartition par modèle constituent une propriété intellectuelle Twingital Institute.

3.3 Pipeline d'inférence

query (SMILES ou nom)

→ Résolution SMILES (RDKit canonicalisation ou PubChem REST)

→ Featurisation moléculaire

→ Encodage, fusion et sélection (propriété intellectuelle Twingital Institute)

→ Calibration probabiliste

→ Évaluation du domaine applicatif

4. Pipeline de données Medallion

4.1 Bronze — Ingestion brute

Source	Volume	Contenu
PubChem	~100 000 composés	SMILES, InChI, MW, descripteurs physicochimiques
ChEMBL v34	~100K composés + activités	IC50, Ki, EC50, données hERG
Tox21	7 831 composés	Labels binaires, 12 assays
Enrichissement NER	~7 300 traités	LD50, LDLo, TDLo, NOAEL, organes

Extraction d'entités nommées toxicologiques par LLM local à partir des monographies PubChem.

4.2 Silver — Curation

Pipeline de curation multi-phases : sanitization, déduplication, filtrage, jointure des labels et des profils enrichis. Résultat : ~145 000 composés dédupliqués.

4.3 Gold — Dataset d'entraînement

Scaffold split Bemis-Murcko strict. Vérification InChIKey : train n val n test = \emptyset .

5. Domaine applicatif

Score AD composite calculé à partir de composantes indépendantes évaluant la similarité structurale, la proximité dans l'espace latent, et la densité régionale.

Les composantes, pondérations, formules de normalisation et seuils de décision constituent une propriété intellectuelle Twingital Institute.

Décision : `is_in_domain` = True/False selon un seuil composite calibré.

6. Calibration probabiliste

Chaque endpoint dispose d'un calibrateur entraîné sur les prédictions out-of-fold du protocole 5-fold CV. ECE post-calibration < 0,05 sur tous les endpoints.

7. Couche d'interprétation LLM

LLM local transformant les scores bruts en rapports toxicologiques pharmacologiquement fondés.

Les mécanismes anti-hallucination, l'architecture du prompt, les paramètres de génération et la structure de la base de connaissances constituent une propriété intellectuelle Twingital Institute.

Dual output : rapport structuré lisible + données structurées machine. Streaming SSE pour masquer la latence.

8. API REST V2.3

8.1 Endpoints

Méthode	Endpoint	Description
POST	/v1/score-toxicity	Prédiction complète 14 endpoints + AD
POST	/v1/interpret	Interprétation LLM
GET	/health	Statut + métriques de validation
GET	/v1/endpoints	Liste détaillée par endpoint

8.2 Codes d'erreur

Code	Code interne	Cause
400	invalid_smiles	SMILES syntaxiquement invalide
400	missing_query	Champ query absent
422	resolution_failed	Nom non résolu via PubChem
422	molecule_too_large	Molécule > 500 atomes lourds
503	model_unavailable	Modèle non chargé ou erreur GPU
504	interpret_timeout	Timeout interprétation LLM

9. Infrastructure

Déploiement sur serveur physique souverain (GPU dédié, pas de dépendance cloud pour le compute). Exposition API via tunnel chiffré. Frontend statique sur CDN. Authentification par magic link. Rate limiting applicatif.

10. Performances détaillées V2.3 (14 endpoints)

Endpoint	Baseline	V2.3 (5-fold)	Cible	Statut
NR-AR	0.584	0.755	0.70	✓
NR-AR-LBD	0.535	0.787	0.70	✓
NR-AhR	0.644	0.902	0.85	✓
NR-Aromatase	0.603	0.887	0.85	✓
NR-ER	0.605	0.854	0.78	✓
NR-ER-LBD	0.645	0.925	0.78	✓
NR-PPAR-γ	0.571	0.752	0.72	✓
SR-ARE	0.630	0.859	0.78	✓
SR-ATAD5	0.687	0.940	0.80	✓
SR-HSE	0.507	0.898	0.78	✓
SR-MMP	0.495	0.909	0.85	✓
SR-p53	0.622	0.934	0.85	✓
Ames Hansen	0.560	0.843	0.87	X (−0.027)
hERG	0.548	0.785	0.83	X (−0.045)
Tox21 MEAN	0.594	0.867	0.80	✓

11. Benchmarks vs littérature

Endpoint	Référence publiée	AUC référence	V2.3	Écart
Ames mutagénicité	DeepAmes (Xu et al. 2021)	0.889	0.843	−0.046
hERG inhibition	CardioTox (Karim et al. 2023)	0.872	0.785	−0.087
NR-AhR	Tox21 challenge winner (2014)	0.911	0.902	−0.009
SR-MMP	AttentiveFP (Xiong et al. 2020)	0.848	0.909	+0.061
NR-Aromatase	GROVER (Rong et al. 2020)	0.882	0.887	+0.005

12. Limites techniques V2.3

Limite	Valeur	Description
Taille max molécule	500 atomes lourds	Au-delà, hors domaine quasi-systématique
Atomes supportés	C H N O S P F Cl Br I	Autres atomes : featurisation dégradée
Métaux de transition	Non supportés	Pt, Ru, Au, etc. : hors périmètre (prévu V3.0)
Peptides	Non recommandé	> 5 acides aminés : hors domaine probable

13. Sécurité et conformité

Positionnement réglementaire : ToxTwin n'est pas un dispositif médical (pas d'usage diagnostique/thérapeutique direct). Scoring toxicologique pré-clinique hors catégorie "haut risque" (EU AI Act, Annexe III). Pipeline conçu par anticipation selon les exigences "haut risque" : audit trail, versionnement modèles, incertitude et domaine applicatif exposés dans chaque réponse API.

14. Plan V2.4 – Contrôle qualité

14.1 Validation interne (réalisée)

5-fold scaffold CV strict, variance inter-fold contrôlée, contamination InChIKey = 0, calibration ECE < 0.05, routing consistant.

14.2 Tests de robustesse (à réaliser)

Reproductibilité inter-runs, invariance des représentations SMILES, sensibilité SAR sur paires connues, couverture du domaine applicatif, monotonie structure-activité.

14.3 Validation externe (à réaliser)

Holdout test gelé (usage unique). Validation prospective externe (accord partenaire ou ECHA). Benchmarks publiés (DeepAmes, CardioTox).

14.4 Axes d'amélioration

Priorité	Action
P1	Enrichissement corpus Ames et hERG
P3	Extension DILI, ClinTox, Carcinogens

Priorité	Action
P3	Support métaux de transition (V3.0)

15. Références

- DiMasi JA et al. *J Health Econ.* 2016;47:20-33.
 - Mayr A et al. DeepTox. *Front Environ Sci.* 2016;3:80.
 - Xiong Z et al. *J Med Chem.* 2020;63(16):8749-8760.
 - Rong Y et al. *NeurIPS.* 2020.
 - Sheridan RP. *JCIM.* 2013;53(4):783-790.
 - Wu Z et al. MoleculeNet. *Chem Sci.* 2018;9(2):513-530.
 - Xu C et al. *Nat. Comm.* 2021.
 - Karim A et al. CardioTox. *JCIM.* 2023.
 - Règlement (UE) 2017/745 (MDR).
 - Règlement (UE) 2024/1689 (EU AI Act).
-