Du système cyber-physique au système stratégique de cyber-résilience

Contrôle spectral, jumeau numérique, intelligence artificielle et mémoire de cas pour renforcer la boucle décisionnelle

Synthèse pédagogique

Août 2026

Introduction — de la réaction automatique à la stratégie

Les systèmes techniques contemporains ne sont plus de simples machines isolées. Un réseau électrique, une installation industrielle, un réseau de télécommunications, un système de transport ou un atelier automatisé forment aujourd'hui des ensembles où la matière, l'énergie, l'information et la décision sont étroitement couplées. Le processus physique produit des mesures ; ces mesures sont transportées par un réseau numérique ; elles sont interprétées par des logiciels ; des décisions sont prises ; enfin des actionneurs modifient le processus physique. La boucle se referme.

C'est précisément ce que l'on appelle un **système cyber-physique** — *Cyber-Physical System*, ou CPS. Le NIST définit les CPS comme des ensembles où interagissent des composants numériques, analogiques, physiques et humains, conçus pour fonctionner par l'intégration de la logique et de la physique. Cette définition est importante : la cybersécurité d'un CPS ne peut pas être réduite à la protection d'un ordinateur ou d'un réseau informatique. Une altération numérique peut modifier un état physique ; inversement, un événement physique peut déclencher une décision numérique. La sécurité doit donc être pensée comme une propriété de la **boucle fermée complète**.

L'idée développée ici consiste à renforcer cette boucle en introduisant deux couches supplémentaires entre la mesure et la commande. La première est une couche de **contrôle spectral de bas niveau** : elle mesure en permanence la marge dynamique qui sépare le système d'une bifurcation ou d'une perte de stabilité. La seconde est une couche plus lente de **délibération**, construite autour d'un jumeau numérique, de la fusion de données, de modèles physiques, d'une mémoire de faits et de cas, et d'une intelligence artificielle capable d'organiser les hypothèses et les scénarios.

La boucle classique

$$capteurs \rightarrow estimation \rightarrow commande$$

se transforme alors en une boucle beaucoup plus riche :

$$capteurs \rightarrow fusion\ et\ estimation \rightarrow jumeau\ numérique \rightarrow diagnostic \rightarrow simulation\ de\ scénarios \rightarrow é$$

Le système ne réagit plus immédiatement à une mesure. Il cherche d'abord à comprendre ce qu'elle signifie, à imaginer plusieurs futurs possibles, à évaluer les conséquences de plusieurs actions et, enfin, à choisir une décision. L'analogie avec les échecs ou le go devient alors féconde : la commande n'est plus seulement un réflexe, elle devient un **coup dans une stratégie**.

1. Anatomie d'un CPS appliqué à un réseau dynamique

1.1 Le processus physique

Considérons un réseau réel représenté par un vecteur d'état

$$X(t) = \left(x_1(t) \ \cdots \ x_n(t)\right)^T.$$

Les composantes peuvent représenter, selon le domaine, des fréquences et puissances électriques, des débits et files d'attente, des températures et pressions, des concentrations biologiques ou des variables collectives d'un réseau complexe.

La dynamique générale s'écrit

$$\dot{X} = F(X, \theta, d, t),$$

où θ représente les paramètres réglables ou inconnus et d les perturbations extérieures. Dans les développements précédents, de nombreux réseaux avaient été ramenés, sous certaines hypothèses, à un Lotka–Volterra généralisé :

$$\dot{X} = X \circ (r - AX).$$

Ici A condense les interactions effectives du réseau et le symbole $\circ$ désigne le produit composante par composante. Cette représentation n'est pas indispensable à la suite : le raisonnement de cyber-résilience vaut pour tout modèle différentiable. Elle est cependant très pratique parce qu'elle rend explicite une matrice d'interaction dont les variations peuvent être estimées et commandées.

1.2 Les capteurs et la mesure

Le processus n'est pas connu directement. Il est observé à travers des capteurs :

$$Y(t) = h(X(t)) + \nu(t),$$

où ν représente les erreurs et le bruit de mesure. Dans le cas le plus simple,

$$Y = X + \nu.$$

L'état utilisé par le calculateur est donc toujours un état **reconstruit**. Cette distinction entre réalité physique et réalité perçue est centrale en cybersécurité : une attaque peut modifier le système lui-même ou seulement la représentation que le contrôleur s'en fait.

1.3 L'estimation et l'identification

À partir des mesures, un observateur construit

$$X\hat{}(t), \theta\hat{}(t), A\hat{}(t).$$

Les techniques possibles vont du filtre de Kalman et de ses variantes à l'identification sous-espace, aux méthodes bayésiennes, aux observateurs robustes et aux méthodes d'apprentissage. Dans notre étude de réseau LV adaptatif, un filtre de Kalman étendu à état augmenté avait été utilisé pour estimer simultanément les trois états et les neuf coefficients d'une matrice $A(t)$ variable.

Cette étape transforme la mesure brute en une **image dynamique interprétable du réseau**.

1.4 Le contrôleur et les actionneurs

La commande $u(t)$ est calculée à partir de l'état estimé :

$$u = C(X\hat{}, A\hat{}, consignes, t).$$

Elle agit sur des paramètres physiques ou numériques : ouverture d'une vanne, réglage d'un convertisseur, redistribution d'une charge, routage d'un trafic, modification d'une priorité, changement d'un gain ou isolement d'un sous-système.

Le CPS est ainsi une boucle où le monde numérique peut modifier le monde physique :

$$X \to Y \to X\hat{} \to u \to X.$$

Cette propriété explique à la fois sa puissance et sa vulnérabilité.

2. La surface de menace : où une cyberattaque peut-elle agir ?

Il est utile de classer les menaces non par type de logiciel malveillant, mais par **point d'injection dans la boucle dynamique**. Cette classification donne immédiatement accès aux conséquences physiques.

2.1 Attaque sur les mesures

Une mesure compromise peut être représentée par

$$Y_a(t) = h(X(t)) + \nu(t) + a_y(t).$$

Le processus physique peut, dans un premier temps, être intact. Mais l'observateur reçoit une information fausse. Il peut reconstruire un mauvais état, une mauvaise matrice d'interaction, puis une mauvaise marge de stabilité.

La chaîne causale est alors

$$attaque\ capteur \rightarrow Y_a \rightarrow X\text{^}erroné \rightarrow A\text{^}erronée \rightarrow mauvaise\ décision$$

Le danger réside dans le fait qu'une commande légitime peut devenir physiquement dangereuse si elle est calculée sur une représentation fausse.

2.2 Attaque sur les actionneurs ou le canal de commande

Le contrôleur calcule u, mais le système reçoit

$$u_a = u + a_u.$$

Cette fois, l'attaque modifie directement la physique. L'algorithme peut croire avoir appliqué une action stabilisante alors qu'une action différente est exécutée. Dans un modèle paramétrique,

$$A(\theta) \rightarrow A\left(\theta + a_\theta\right).$$

Le spectre de la dynamique réelle se déplace alors sans que le calculateur en ait nécessairement conscience.

2.3 Attaque sur un paramètre ou une configuration du réseau

Une modification de configuration peut être schématisée par

$$A \rightarrow A + \Delta A_{cyber}.$$

Dans un réseau réel, les coefficients de A peuvent représenter des couplages, gains, priorités, capacités ou poids de routage. Une petite modification d'un coefficient particulièrement sensible peut avoir un effet dynamique disproportionné.

Pour un GLV, l'équilibre intérieur vérifie

$$X^* = A^{-1}r,$$

et le Jacobien à l'équilibre est

$$J^* = -\ diag\left(X^*\right)A.$$

Une altération de A entraîne donc simultanément un déplacement de l'équilibre et du spectre :

$$A \rightarrow X^* \rightarrow J^* \rightarrow \lambda_i\left(J^*\right).$$

Une attaque peut ainsi ne provoquer aucune panne immédiate tout en réduisant progressivement la marge dynamique du réseau.

2.4 Attaque sur la topologie et les communications

La suppression, le retard ou la modification d'un lien change l'information effectivement disponible et, parfois, la structure physique du réseau. On peut représenter cette situation par une modification d'une matrice d'adjacence, d'un Laplacien ou d'un opérateur de couplage.

La conséquence peut être double : perte de connectivité et perte de stabilité dynamique. Il faut donc distinguer la **robustesse topologique**, liée par exemple au spectre d'un Laplacien, et la **résilience dynamique**, liée au Jacobien de la dynamique fermée.

2.5 Attaque coordonnée et furtive

Le scénario le plus délicat combine plusieurs points d'injection. L'attaque modifie réellement la dynamique,

$$A \to A + \Delta A,$$

mais altère simultanément les mesures pour masquer l'effet :

$$Y \to Y + a_y.$$

Il devient alors possible, en principe, que la marge réelle se dégrade tandis que la marge estimée semble normale :

$$\rho_{vrai} \ll \rho\hat{}.$$

C'est précisément pour ce type de situation qu'un simple observateur n'est plus suffisant. Il faut confronter plusieurs sources d'information, plusieurs modèles et plusieurs hypothèses.

2.6 Le jumeau numérique et l'IA deviennent eux-mêmes des cibles

Lorsqu'une couche de simulation et de décision est ajoutée, elle agrandit aussi la surface d'attaque. Les flux qui synchronisent le jumeau peuvent être corrompus ; le modèle peut devenir obsolète ; une base de connaissances peut être altérée ; une IA peut être trompée ou mal interpréter une situation. Le NIST IR 8356 insiste précisément sur les enjeux de confiance, d'intégrité des données, de synchronisation, de contrôle et de sécurité propres aux jumeaux numériques.

Le principe de défense doit donc être récursif : **le système qui surveille doit lui-même être surveillé**.

3. Des conséquences informatiques aux conséquences dynamiques

La gravité d'une attaque ne devrait pas être évaluée uniquement à partir de sa sophistication informatique. La question essentielle est : *que fait-elle à la dynamique physique du système ?*

Une attaque peut modifier un paramètre très peu influent, auquel cas l'impact physique est limité. Une autre peut modifier très légèrement un paramètre situé dans une direction critique du modèle, et rapprocher fortement le réseau d'une bifurcation.

Cette distinction conduit à construire une mesure d'impact fondée sur la dynamique.

4. La couche basse : définir et contrôler une résilience spectrale

4.1 Linéarisation locale

Autour d'un état de fonctionnement X^*, on écrit

$$\delta X^{\cdot} = J\delta X,$$

où

$$J = \frac{\partial F}{\partial X})_{X^*}.$$

Les valeurs propres de J déterminent la réponse asymptotique des petites perturbations. Si toutes leurs parties réelles sont négatives, le régime est localement asymptotiquement stable.

4.2 Marge de résilience spectrale

On définit l'abscisse spectrale

$$\alpha(J) = \max_i \Re[\lambda_i(J)).$$

Puis une marge de stabilité

$$\rho = -\alpha(J).$$

Pour un système localement stable,

$$\rho > 0.$$

Lorsque

$$\rho \to 0, \rho > 0,$$

la valeur propre dominante approche la frontière de stabilité et le temps de relaxation augmente. Une approximation naturelle est

$$\tau_{retour} \sim \frac{1}{\rho}.$$

Dans ce sens précis, ρ peut être interprété comme une **résilience locale asymptotique** : elle mesure la rapidité avec laquelle le système récupère d'une petite perturbation.

4.3 Une attaque comme déplacement de la marge

Si une attaque produit ΔA, la variation de résilience peut être développée au premier ordre :

$$\Delta\rho \simeq \sum_{i,j} \frac{\partial \rho}{\partial a_{ij}} \Delta a_{ij}.$$

On définit ainsi une sensibilité dynamique

$$S_{ij}^{(\rho)} = \frac{\partial \rho}{\partial a_{ij}}.$$

Sa valeur absolue donne une mesure de criticité : un paramètre pour lequel $|S_{ij}^{(\rho)})$ est élevé mérite une protection et une surveillance renforcées.

4.4 Contrôle spectral

Supposons que le système puisse agir sur certains paramètres θ_p :

$$A(\theta) = A_0 + \sum_{p=1}^{m} \theta_p B_p.$$

On fixe une consigne de résilience

$$\rho_c > 0,$$

et l'erreur

$$\varepsilon_\rho = \rho\hat{} - \rho_c.$$

Une loi adaptative locale simple est

$$\dot{\theta} = -\Gamma\left(\rho\hat{} - \rho_c\right)\nabla_\theta \rho\hat{}.$$

Cette formulation ne doit pas être vue comme la solution industrielle finale, mais comme une preuve de concept : **la marge de stabilité devient une variable asservissable**.

4.5 Au-delà des valeurs propres : non-normalité et robustesse

Une réserve est essentielle. Pour une matrice non normale, toutes les valeurs propres peuvent être stables tandis qu'une perturbation connaît une forte amplification transitoire. Un second indicateur utile est l'abscisse numérique

$$\eta(J) = \lambda_{max}\left(\frac{J+J^T}{2}\right).$$

Une architecture robuste ne devrait donc pas réduire la cyber-résilience à un seul scalaire. On peut envisager un vecteur de sûreté

$$R = \left(\rho, -\ \eta, d_{\partial B}, \sigma_{\rho}\right),$$

où $d_{\partial B}$ représente une distance à la frontière du bassin d'attraction et σ_{ρ} une incertitude sur l'estimation spectrale.

4.6 Fonction du contrôle spectral dans l'architecture complète

Cette couche doit rester rapide, déterministe, transparente et vérifiable. Elle joue le rôle d'un **réflexe de stabilité**. Elle peut imposer des enveloppes de sûreté au niveau supérieur :

$$\rho \geq \rho_{min}, X \in D_{sûr}, u \in U_{admissible}.$$

La couche stratégique peut ensuite optimiser la mission à l'intérieur de cette enveloppe.

5. Pourquoi intercaler un jumeau numérique entre les capteurs et la commande ?

Le jumeau numérique introduit une capacité nouvelle : **essayer sans agir**. Il reçoit les données du système, les fusionne avec un modèle et peut simuler les conséquences de plusieurs hypothèses ou commandes avant de toucher au système réel.

5.1 Le jumeau comme observateur augmenté

Le jumeau maintient un état interne

$$Z_{DT}(t) = \{X\hat{}, A\hat{}, J\hat{}, \rho\hat{}, topologie, incertitudes, contexte\}.$$

Il peut produire un état prédit

$$X\hat{}(t\ +\ \Delta t|t)$$

et comparer cette prédiction à la nouvelle mesure. Le résidu

$$e(t) = Y(t) - Y\hat{}(t|t\ -\ \Delta t)$$

constitue un premier indicateur d'incohérence.

Le travail de Balta et collaborateurs au NIST et à l'Université du Michigan illustre cette idée dans un cas expérimental de fabrication additive : un jumeau utilisant modèles physiques, modèles de données et connaissance experte est employé pour distinguer des anomalies attendues de cyberattaques pendant des transitoires contrôlés.

5.2 Le jumeau comme laboratoire de diagnostic

Une anomalie observée peut avoir plusieurs causes. Le jumeau peut tester :

- H_1 : fluctuation normale ;
- H_2 : panne physique ;
- H_3 : capteur défaillant ou compromis ;
- H_4 : actionneur compromis ;
- H_5 : modification d'un paramètre ;
- H_6 : attaque coordonnée.

Pour chaque hypothèse, il simule une trajectoire

$$X_{H_i}(t + \tau)$$

et mesure sa compatibilité avec les données. Le diagnostic devient un **raisonnement par modèles concurrents**.

5.3 Le jumeau comme laboratoire de contre-mesures

Une fois plusieurs explications plausibles identifiées, on teste différentes réponses u_j :

$$X_{ij}(t + \tau) = \Phi_{H_i}\left(X(t), u_j, \tau\right).$$

On évalue alors chaque combinaison hypothèse-action selon plusieurs critères : respect de la mission, coût, sûreté, cyber-risque, temps de récupération et résilience spectrale.

Le jumeau devient ainsi une couche de **délibération contrefactuelle** : *si le monde est dans cet état et si j'applique cette action, que se passe-t-il ?*

6. La fusion de données : reconstruire une « position » fiable

Un joueur d'échecs ne peut raisonner que s'il connaît correctement la position des pièces. Un CPS doit, lui aussi, construire une représentation cohérente à partir de sources hétérogènes.

La fusion de données peut inclure :

- télémétrie physique ;
- journaux de communication ;
- événements des contrôleurs ;
- alarmes cyber ;
- topologie ;
- historique de maintenance ;
- configuration logicielle ;
- informations humaines ou opérationnelles.

La revue de Wang et collaborateurs sur la fusion multi-source dans les Cyber-Physical-Social Systems montre l'importance croissante de la combinaison entre apprentissage profond et graphes de connaissances : l'un absorbe de grands flux de données, l'autre apporte structure, relations, contexte et possibilité d'inférence.

Cette fusion peut être formulée comme l'estimation d'un état de croyance

$$b_t = P\left(Z_t | D_{0:t}\right),$$

où $D_{0:t}$ regroupe les données disponibles jusqu'à l'instant t.

Le système stratégique ne décide donc pas sur « la vérité », mais sur une **distribution d'hypothèses accompagnée d'incertitudes**.

7. La couche IA : non pas remplacer la physique, mais organiser le raisonnement

L'IA est particulièrement utile au-dessus du jumeau, là où le problème devient combinatoire et sémantique. Le modèle physique répond bien à la question : *que se passerait-il si ?* L'IA aide à répondre aux questions : *quelles hypothèses valent la peine d'être testées ? quels scénarios sont plausibles ? quels faits antérieurs ressemblent à la situation actuelle ?*

7.1 Une architecture hybride

Une architecture robuste peut associer quatre formes de connaissance :

$$physique + statistiques + règles\ expertes + apprentissage$$

Le modèle physique impose les invariants et les limites de sûreté. Les statistiques quantifient l'incertitude. Les règles expertes codent les procédures métier. L'apprentissage reconnaît des motifs et propose des hypothèses dans les cas où la modélisation explicite est incomplète.

Les travaux 2026 de HySecTwin illustrent cette orientation : un jumeau numérique de CPS est enrichi par une représentation sémantique et un moteur de raisonnement combinant règles déterministes et raisonnement flou afin de produire des évaluations interprétables et auditables à partir de télémétrie temps réel. Il s'agit encore d'un travail de recherche récent, mais il est très proche conceptuellement de la couche experte envisagée ici.

7.2 IA contrainte par la physique

L'usage d'un grand modèle de langage dans une boucle critique ne peut pas être celui d'un agent libre. Le travail 2026 de Kampourakis, Gkioulos et Katsikas sur le banc SWaT adopte une architecture intéressante : les heuristiques déterministes traitent les signatures connues ; le LLM n'est invoqué que dans les cas ambigus ; ses sorties sont contraintes par un schéma et filtrées par des tests de plausibilité physique.

Le principe général est :

$$IA\ propose ; modèle\ physique\ vérifie ; règles\ de\ sûreté\ autorisent\ ou\ refusent.$$

7.3 Une mémoire de faits, de cas et de conséquences

La stratégie s'améliore si le système possède une mémoire structurée. Elle peut être divisée en trois niveaux.

Mémoire factuelle : topologie, équipements, versions, contraintes, paramètres, incidents connus.

Mémoire épisodique : événements passés avec leurs chronologies, diagnostics, actions et résultats.

Mémoire de règles : procédures, politiques de sûreté, contraintes réglementaires, règles métier.

Une entrée de mémoire de cas peut prendre la forme :

$$C_k = (contexte, symptômes, hypothèses, action, résultat, \Delta\rho).$$

Lors d'un nouvel événement, l'IA peut rechercher les cas proches, non pour les recopier, mais pour réduire l'espace des hypothèses et proposer les simulations les plus informatives.

Les graphes de connaissances constituent une technologie particulièrement adaptée à cette mémoire relationnelle. La revue 2024 de Listl et collaborateurs montre leur intérêt pour relier des modèles de simulation hétérogènes, automatiser l'adaptation de jumeaux et organiser les connaissances industrielles.

8. De l'automatisme au jeu d'échecs : construire une stratégie

8.1 Le « coup » comme commande

À l'instant t, supposons que le système dispose d'un ensemble d'actions admissibles

$$U(t) = \{u_1, ..., u_b\}.$$

Chaque action produit un futur différent dans le jumeau. Si l'on regarde d décisions à l'avance, l'arbre brut contient de l'ordre de

$$N_{branches} \sim b^d$$

séquences. Avec seulement vingt choix par étape et dix étapes d'horizon, on atteint déjà environ 10^{13} séquences. L'exploration exhaustive est donc impossible dans un système réel.

8.2 Pourquoi l'analogie avec les échecs et le go est utile

Les bons moteurs de jeu ne calculent pas « tous les coups jusqu'à la fin ». Ils combinent :

- une représentation compacte de l'état ;
- une génération de coups légaux ;
- une fonction d'évaluation ;
- un élagage des branches peu prometteuses ;
- une recherche à horizon limité ;
- une mémoire de positions ou d'expériences ;
- éventuellement une fonction apprise qui guide la recherche.

Un système de décision CPS peut fonctionner de la même façon.

8.3 Fonction d'évaluation

Une séquence de commandes U peut être évaluée par

$$J(U) = w_1 C_{mission} + w_2 C_{énergie} + w_3 C_{cyber} + w_4 C_{indisponibilité} - w_5 \rho_{future}.$$

On cherche

$$U^* = arg \min_U J(U)$$

sous les contraintes de sûreté.

La résilience spectrale n'est donc pas toute la fonction d'évaluation ; elle en est une composante physique transparente, particulièrement utile pour éviter qu'une stratégie apparemment performante rapproche le système d'un seuil dynamique.

8.4 Model Predictive Control : une première réalisation pratique de cette idée

Le MPC est déjà une forme de « jeu à plusieurs coups ». À chaque instant, il prédit l'évolution sur un horizon, optimise une séquence de commandes, applique seulement la première, puis recommence avec les nouvelles mesures.

$$mesurer \rightarrow prédire \rightarrow optimiser \rightarrow agir\ une\ fois \rightarrow remesurer$$

Les travaux de Chen et collaborateurs en 2025 combinent jumeau numérique, prédiction multi-pas par réseau neuronal temporel et MPC pour la fabrication additive. Ils cherchent explicitement à accélérer la prédiction de tout l'horizon afin de permettre une décision temps réel. Un prolongement du même groupe intègre l'incertitude par prédiction quantile et MPC robuste.

8.5 Lorsque l'environnement contient un adversaire

En cybersécurité, une action du défenseur peut provoquer une réponse de l'attaquant. Le problème prend une forme de jeu :

$$\min_{u} \max_{a} J(u, a).$$

Cette écriture ne signifie pas que l'on connaît toutes les actions possibles de l'adversaire. Elle exprime le principe de robustesse : choisir une action qui reste acceptable face à des réponses défavorables plausibles.

Si l'on associe des probabilités aux hypothèses d'attaque, on peut utiliser une espérance de coût

$$E[J(u, a))$$

ou un critère de risque focalisé sur la queue de distribution, par exemple un CVaR. La stratégie devient alors une combinaison de théorie des jeux, contrôle robuste, décision sous incertitude et simulation.

9. Une architecture hiérarchique proposée

L'ensemble des éléments précédents peut être organisé en six couches.

Couche 1 — processus, capteurs et actionneurs

Le monde physique évolue. Les capteurs produisent des mesures. Les actionneurs exécutent des commandes.

Couche 2 — fusion et estimation résiliente

Cette couche combine plusieurs sources et estime l'état, les paramètres et les incertitudes. Elle surveille également les résidus et les incohérences. La littérature sur l'estimation résiliente de CPS montre depuis plusieurs années comment reconstruire l'état malgré des capteurs ou actionneurs compromis ; des travaux sur l'IEEE 68-bus ont, par exemple, illustré des algorithmes de détection, identification et mitigation de falsifications de données.

Couche 3 — contrôle spectral de sûreté

Elle calcule rapidement

$$\rho\hat{} =- max\Re\lambda(J\hat{})$$

et impose une enveloppe :

$$\rho\hat{} \geq \rho_{min}.$$

Cette couche doit pouvoir fonctionner même si la stratégie supérieure est indisponible. Elle constitue un **garde-fou local et rapide**.

Couche 4 — jumeau numérique

Le jumeau synchronise données, modèles et simulation. Il teste les hypothèses et les commandes possibles et quantifie leurs conséquences.

Couche 5 — intelligence stratégique et mémoire

Cette couche gère les hypothèses, sélectionne les branches à simuler, consulte la mémoire de cas, applique les règles expertes et construit une stratégie candidate.

Les architectures « agentic digital twin » apparues dans la littérature 2026 cherchent précisément à faire évoluer le jumeau de la prédiction vers le raisonnement et la prescription. AgenticTwin, publié en prépublication en août 2026, associe par exemple un pipeline d'anomalie fondé sur un jumeau à une collaboration structurée entre agents LLM afin d'améliorer diagnostic, recherche contextuelle et proposition de mitigation. Ces résultats sont très récents et doivent être considérés comme exploratoires, mais ils illustrent la direction générale.

Couche 6 — gouvernance et humain dans la boucle

Pour les infrastructures critiques, toutes les décisions ne doivent pas avoir le même niveau d'autonomie. On peut distinguer :

1. recommandations sans action automatique ;
2. actions automatiques sûres, limitées et réversibles ;
3. actions importantes soumises à validation humaine ;
4. procédures d'urgence déterministes indépendantes de l'IA.

Le NIST AI RMF insiste sur la nécessité de définir explicitement les rôles humains et les niveaux d'autonomie. L'architecture NIST de jumeau numérique *human-in-the-loop* va dans le même sens en intégrant espaces physique, cyber et humain pour soutenir la décision collaborative.

10. Exemple narratif : une dégradation progressive d'un réseau

Imaginons un réseau en régime stable. Sa marge spectrale nominale vaut

$$\rho = 0.15.$$

Une dérive de configuration apparaît sur un sous-système. Les variables de service restent dans leurs plages habituelles, mais le jumeau estime

$$\rho = 0.09.$$

La couche spectrale ne déclenche pas encore d'action d'urgence, mais signale une perte de marge. La fusion de données détecte simultanément une modification inhabituelle de configuration et quelques résidus capteurs.

Le système stratégique construit alors plusieurs hypothèses : défaut de composant, erreur de maintenance, capteur défaillant, modification malveillante. Il sélectionne les simulations susceptibles de mieux différencier ces scénarios. Pour chaque hypothèse, plusieurs actions sont essayées dans le jumeau : ne rien faire, redistribuer une charge, isoler un lien, basculer sur une redondance, renforcer temporairement un paramètre stabilisant.

Une action permet de restaurer

$$\rho \simeq 0.13$$

avec un coût opérationnel faible, mais elle serait dangereuse si l'hypothèse « capteur falsifié » était vraie. Le système consulte alors la mémoire : un événement proche avait été résolu par recoupement avec un capteur indépendant. Il propose donc d'abord une mesure de discrimination peu coûteuse, puis replanifie à partir du résultat.

Ce comportement est beaucoup plus proche d'une stratégie que d'un automatisme. Le système cherche un **coup informatif** avant un coup correctif.

11. Limites et questions de recherche

11.1 Fidélité du jumeau

Un jumeau n'est jamais le système réel. Il doit être synchronisé, calibré et accompagné d'une quantification de l'incertitude. Une stratégie optimale dans un mauvais modèle peut être dangereuse.

11.2 Explosion combinatoire

L'arbre des scénarios croît exponentiellement. Il faut donc des méthodes d'élagage, des modèles réduits, des prédicteurs rapides et des politiques apprises. Le défi est d'accélérer le calcul sans perdre les garanties de sûreté.

11.3 Identifiabilité et observabilité

Une attaque peut exploiter les directions mal observées. L'estimation de $A(t)$ ou de $J(t)$ exige une excitation suffisante et des capteurs suffisamment informatifs. La fusion de données ne remplace pas l'observabilité mathématique.

11.4 Incertitude et changement de régime

Le modèle peut devenir faux à la suite d'un changement structurel. L'IA et l'observateur doivent distinguer dérive normale, rupture de modèle et attaque.

11.5 Sécurité de la mémoire et de l'IA

Une mémoire de faits devient un actif critique. Son empoisonnement peut biaiser les stratégies futures. Les sources doivent être authentifiées et les connaissances versionnées et auditables.

11.6 Explicabilité

La commande d'une infrastructure critique doit pouvoir être expliquée. Une recommandation devrait exposer : faits utilisés, hypothèses, incertitudes, scénarios simulés, contraintes vérifiées, effet prévu sur la mission et sur la marge de résilience.

11.7 Résilience locale versus résilience globale

La marge spectrale est locale. Elle ne donne pas à elle seule la taille du bassin d'attraction ni la capacité à survivre à une perturbation de grande amplitude. Elle doit être complétée par des simulations non linéaires et des mesures de robustesse globale.

12. Conclusion — vers un système qui « pense avant d'agir »

Le CPS classique est déjà une boucle décisionnelle : il mesure, calcule et agit. Le jumeau numérique permet d'ajouter une étape essentielle : **simuler avant d'agir**. La fusion de données permet de reconstruire une position plus fiable ; le contrôle spectral fournit une mesure rapide de distance aux seuils dynamiques ; l'IA et la mémoire de cas permettent d'organiser la complexité des hypothèses et des futurs possibles.

L'analogie avec les échecs ou le go ne signifie pas que le réseau doit explorer tous les futurs. Elle signifie que l'on peut organiser la décision selon les mêmes principes : construire une position, générer des coups admissibles, éliminer les branches mauvaises, regarder plusieurs coups à l'avance, utiliser une fonction d'évaluation et apprendre de l'expérience.

La différence essentielle est que les règles du jeu ne sont pas purement symboliques. Elles sont imposées par la physique. C'est pourquoi le jumeau numérique et le modèle dynamique doivent rester au centre du raisonnement.

L'architecture recherchée pourrait finalement être résumée ainsi :

$$percevoir \rightarrow estimer \rightarrow comprendre \rightarrow imaginer \rightarrow évaluer \rightarrow décider \rightarrow agir \rightarrow apprendre$$

Le contrôle spectral constitue le garde-fou dynamique ; le jumeau numérique constitue le monde dans lequel les futurs sont essayés ; l'IA constitue l'outil qui organise les hypothèses et la recherche ; la mémoire transforme les expériences passées en connaissances réutilisables ; l'humain conserve la gouvernance des décisions qui engagent fortement la sûreté. La cyber-résilience devient alors moins une propriété passive qu'une **capacité active à comprendre, anticiper, absorber et reconfigurer**.

Bibliographie commentée — fondements, cas d'application et recherches récentes

A. Fondements CPS et gouvernance

Griffor, E. R., Greer, C., Wollman, D. A., & Burns, M. J. (2017). *Framework for Cyber-Physical Systems: Volume 1, Overview*. NIST SP 1500-201. DOI: 10.6028/NIST.SP.1500-201.
Référence de base pour la définition d'un CPS comme ensemble intégré de composants numériques, analogiques, physiques et humains. Le cadre insiste sur la confiance, le temps, les données, la composition et le cycle de vie. Il fournit la charpente conceptuelle dans laquelle s'inscrit l'architecture proposée.

NIST (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1.
Le cadre NIST n'est pas spécifique aux jumeaux numériques, mais il est particulièrement pertinent pour la couche IA. Il insiste sur la gouvernance, la mesure du risque, la sécurité, la résilience, la transparence et la définition explicite des rôles de l'humain et de l'automatisation. Il justifie une architecture où les niveaux d'autonomie sont différenciés.

B. Jumeaux numériques et cybersécurité

Balta, E., Pease, M., Moyne, J., Barton, K., & Tilbury, D. (2023). "Digital Twin-Based Cyber-Attack Detection Framework for Cyber-Physical Manufacturing Systems." *IEEE Transactions on Automation Science and Engineering*. DOI: 10.1109/TASE.2023.3243147.
Cas d'application important : le jumeau est utilisé sur des imprimantes 3D commerciales pour distinguer anomalies attendues et cyberattaques pendant des transitoires contrôlés. Le cadre combine modèles physiques, modèles de données et expertise métier. C'est l'un des exemples les plus proches de la fonction « comparaison réel–jumeau → diagnostic » décrite dans ce document.

Voas, J., Mell, P., Laplante, P., & Piroumian, V. (2025). *Security and Trust Considerations for Digital Twin Technology*. NIST IR 8356. DOI: 10.6028/NIST.IR.8356.
Référence institutionnelle centrale. Elle rappelle que le jumeau numérique apporte de nouvelles fonctions de surveillance, simulation et commande, mais crée aussi de nouveaux problèmes de confiance : intégrité des données, synchronisation, sécurité, instrumentation et contrôle. Indispensable pour éviter de traiter le jumeau comme un observateur intrinsèquement fiable.

Kim, D., Bajestani, M., Shao, G., Jones, A. T., & Noh, S. D. (2023). "Conceptual Architecture of Digital Twins with Human-in-the-Loop-Based Smart Manufacturing." IMECE 2023.
Propose une architecture intégrant espaces physique, cyber et humain. Le papier est particulièrement pertinent pour la couche de gouvernance : l'IA et la cognition humaine sont utilisées conjointement dans la décision. Un cas de fabrication additive métallique illustre l'architecture.

C. Estimation résiliente des CPS

Yong, S. Z., Zhu, M., & Frazzoli, E. (2017). "Switching and Data Injection Attacks on Stochastic Cyber-Physical Systems: Modeling, Resilient Estimation and Attack Mitigation."
Travail important sur l'estimation d'état malgré des attaques de commutation et de falsification de données sur capteurs et actionneurs. Les auteurs discutent détection, identification et mitigation, avec des simulations incluant un réseau IEEE 68-bus. Ce type de travail fonde la couche d'estimation résiliente située sous le jumeau stratégique.

Fawzi, H., Tabuada, P., & Diggavi, S. (2014, travaux initiés en 2012). "Secure Estimation and Control for Cyber-Physical Systems under Adversarial Attacks."
Référence fondatrice sur la capacité à reconstruire un état et à stabiliser un système lorsque certains capteurs ou actionneurs sont compromis. Elle montre que la résilience du contrôle dépend d'abord de la possibilité d'une estimation fiable.

D. Fusion de données, sémantique et mémoire

Wang, X., Wang, Y., Yang, J., Jia, X., Li, L., Ding, W., & Wang, F.-Y. (2024). "The Survey on Multi-Source Data Fusion in Cyber-Physical-Social Systems." arXiv:2404.07476.
Revue récente sur la fusion multi-source. Elle propose notamment de combiner apprentissage profond et graphes de connaissances afin d'obtenir à la fois capacité de traitement à grande échelle et meilleure structuration sémantique. Très pertinente pour la construction de la « position » utilisée par le système stratégique.

Listl, F. G., Dittler, D., Hildebrandt, G., Stegmaier, V., Jazdi, N., & Weyrich, M. (2024). "Knowledge Graphs in the Digital Twin: A Systematic Literature Review About the Combination of Semantic Technologies and Simulation in Industrial Automation." arXiv:2406.09042.
Montre comment les graphes de connaissances peuvent relier des sources hétérogènes et des modèles de simulation dans les jumeaux numériques. Cette approche est directement pertinente pour une mémoire structurée de faits, relations, règles et cas.

E. Décision multi-scénarios et MPC

Chen, Y.-P., Karkaria, V., Tsai, Y.-K., Rolark, F., Quispe, D., Gao, R. X., Cao, J., & Chen, W. (2025). "Real-Time Decision-Making for Digital Twin in Additive Manufacturing with Model Predictive Control using Time-Series Deep Neural Networks."
Cas d'application particulièrement instructif : un prédicteur multi-pas fournit tout l'horizon au MPC afin d'accélérer la décision temps réel dans un procédé de dépôt de matière par énergie dirigée. Le travail illustre concrètement l'idée « simuler plusieurs étapes avant d'appliquer la première commande ».

Chen, Y.-P., Tsai, Y.-K., Karkaria, V., & Chen, W. (2025–2026). "Uncertainty-Aware Digital Twins: Robust Model Predictive Control using Time-Series Deep Quantile Learning."
Prolonge la logique précédente en ajoutant une quantification d'incertitude par régression quantile. L'objectif est de conserver une marge de sûreté face aux perturbations et à

l'incertitude du modèle. Cette approche est très proche de la notion de stratégie sous incertitude développée dans ce texte.

F. Raisonnement hybride et IA pour jumeaux cyber-physiques — travaux 2026

Holmes, D., Moshin, A., Nepal, S., Sikos, L., & Yanicke, H. (2026). "HySecTwin: A Knowledge-Driven Digital Twin Framework Augmented with Hybrid Reasoning for Cyber-Physical Systems." arXiv:2605.11682.
Travail récent et encore émergent. Il place un moteur de raisonnement au coeur du jumeau, avec modélisation sémantique, règles déterministes et raisonnement flou. Les résultats annoncés sur un banc CPS montrent l'intérêt de la sémantique pour produire des évaluations interprétables et auditables. C'est probablement l'une des références les plus proches de l'idée de « système expert au-dessus du jumeau ».

Kampourakis, K. E., Gkioulos, V., & Katsikas, S. (2026). "Systematic Integration of Digital Twins and Constrained LLMs for Interpretable Cyber-Physical Anomaly Detection." arXiv:2604.03790.
Architecture hybride évaluée sur le procédé industriel SWaT. Des heuristiques identifient des signatures connues et le LLM n'intervient que lorsque les heuristiques s'abstiennent. Les sorties sont contraintes par un format structuré et des filtres de plausibilité. Cette logique de LLM limité par la physique est particulièrement pertinente pour les CPS critiques.

Hasan, T., Ghanta, M., Sarkar, S., & Guin, U. (2026). "AgenticTwin: An Agentic LLM Framework Integrated with Digital Twin for Anomaly Detection." arXiv:2608.11679.
Prépublication très récente (août 2026). Elle associe un pipeline de détection fondé sur un jumeau numérique à plusieurs agents LLM capables d'expliquer, rechercher du contexte et proposer des mitigations. Les auteurs évaluent également des modèles légers déployables localement. Cette publication doit être considérée comme un indicateur de direction de recherche, non comme une technologie déjà industrialisée.

G. Perspectives : « agentic digital twins » et CPS autonomes

Hasan et al. (2026), travaux de synthèse sur l'intégration d'agentic AI et de digital twins pour les infrastructures cyber-physiques.
La littérature 2026 commence à formaliser le passage du jumeau réactif au jumeau prescriptif ou agentique : le jumeau n'est plus seulement un modèle synchronisé, il devient une plateforme de planification et d'orchestration. Ce champ reste jeune ; les questions de validation, de sûreté et de responsabilité y sont centrales.

Recherches 2026 sur la sécurité des agents IA dans les CPS.
Les travaux récents insistent sur le fait que l'agent doit être évalué en boucle fermée : délais, état périmé, actions irréversibles, couplage avec l'infrastructure et compromission des outils modifient profondément le risque. Le jumeau est alors utile comme environnement de validation avant déploiement réel.

Repères de lecture

Pour une première lecture centrée sur l'architecture proposée, l'ordre suivant est conseillé :

1. **Griffor et al. (NIST, 2017)** pour la définition du CPS ;
2. **Balta et al. (2023)** pour la détection cyber par jumeau numérique ;
3. **Voas et al. (NIST IR 8356, 2025)** pour la confiance et la sécurité du jumeau ;
4. **Chen et al. (2025)** pour la décision multi-pas par MPC ;
5. **Wang et al. (2024)** et **Listl et al. (2024)** pour la fusion et la mémoire sémantique ;
6. **Holmes et al. (2026)** pour le raisonnement hybride ;
7. **Kampourakis et al. (2026)** pour l'IA générative contrainte par la physique ;
8. **AgenticTwin (2026)** pour la direction « jumeau + agents + mémoire + mitigation ».

Note méthodologique. Les références NIST et les articles évalués par les pairs constituent les sources les plus solides du corpus. Plusieurs références de 2026 sont des prépublications récentes : elles sont utilisées ici pour documenter les tendances de recherche les plus actuelles, et non comme preuves d'une maturité industrielle déjà acquise.