



Pour le déploiement d'une industrie de production
numérisée, résiliente et éthique en Nouvelle-Aquitaine

Charte IA digne de confiance

version 2025



Cofinancé par
l'Union européenne



RÉGION
Nouvelle-
Aquitaine

Projet cofinancé par la
Région Nouvelle-Aquitaine

Introduction

Pour une innovation en Intelligence Artificielle digne de confiance

> Un contexte en Intelligence Artificielle et innovation des plus dynamiques

L'adoption de l'**AI Act** en 2024 a marqué un tournant dans le domaine de l'Intelligence Artificielle¹ au niveau mondial et européen.

L'adoption de ce **cadre juridique** historique pour encadrer le développement et l'utilisation de l'**Intelligence Artificielle (IA) et des systèmes d'IA (SIA) de confiance** en Europe illustre une volonté de rassurer et de fédérer autour d'une vision commune les acteurs d'un domaine dont les récentes prouesses étaient encore inimaginables il y a 2 ans.

Ces acteurs : des décideurs politiques confrontés à la nécessité de comprendre pour légiférer et décider, des chefs d'entreprise dont les services et les revenus sont challengés par ces nouveaux outils, des salariés dont les métiers sont impactés, des chercheurs au cœur des innovations et le grand public, consommateur final de produits et d'informations générés ou non par IA.

Cette technologie, représentant un fort potentiel de croissance, possède également un fort potentiel de perturbation des activités et du quotidien des entreprises et autorités publiques, de l'individu aux structures en elles-mêmes.

En cause, **des défis éthiques** complexes présents depuis les premières phases d'idéation, à la conception, au déploiement de systèmes à base d'IA, parmi lesquels – sans être exhaustif : ceux de la traçabilité et de la gouvernance des données, de la cybersécurité, des hallucinations et des comportements biaisés des SIA, du manque de transparence et d'explicabilité, de la manipulation à des fins malveillantes et de la désinformation.

L'AI Act s'est ainsi imposé comme un cadre légal structurant pour proposer une innovation, en IA et avec l'IA, désormais responsable, transparente et durable.

> Dihnamic, acteur de l'innovation de confiance porté par la région Nouvelle-Aquitaine et l'Europe

Dans ce contexte, **Dihnamic** (Digital Innovation Hub for Nouvelle-Aquitaine Manufacturing Industry Community) **accompagne l'innovation des autorités publiques et des entreprises manufacturières et entreprises de service à l'industrie depuis 2022**. Avec une expertise autour de 4 nœuds technologiques – IA, Internet des Objets, Robotique et Jumeaux numériques - Dihnamic promeut une **IA éthique** et un **numérique responsable** grâce à des services et actions en lien avec l'axe « IA digne de confiance ».

Il faut comprendre ici « confiance en IA » et « IA éthique » comme étant des propriétés émergentes issues d'un système dans un contexte métier, applicatif et lié à un niveau de risque. Ces propriétés sont à caractériser dès les premières phases des projets d'innovation en s'aidant d'une stratégie d'innovation pluridisciplinaire, car les approches techniques seules sont insuffisantes pour atteindre un niveau de confiance cohérent et adéquat selon le contexte métier et le niveau de risque associé.

¹ Le glossaire figurant à la fin du présent document fournit les définitions discutées par le hub d'experts de "l'Intelligence Artificielle" et du "Numérique responsable" aux fins de ce même document



> Notre charte pour sensibiliser à une IA digne de confiance

Pour accompagner cette étape désormais incontournable des projets IA, Dihnamic propose **une charte pour sensibiliser à une IA digne de confiance** présentant des recommandations simplifiées mais fortes autour de valeurs essentielles, telles que le contrôle humain, la sécurité, la gouvernance des données, la transparence, le développement durable et l'équité.

Cette charte est issue :

- de l'accompagnement d'un Hub d'experts spécialisés en IA et en éthique du numérique,
- d'échanges avec plus de 200 entreprises et partenaires de l'écosystème de Nouvelle-Aquitaine,
- de l'étude de nombreux travaux et productions à la fois méthodologiques et techniques open-source des communautés techniques et scientifiques autour de l'IA digne de confiance et des thématiques attenantes,
- des travaux de programmes régionaux², de grands programmes nationaux portés par les acteurs publics français et européen autour du Numérique Responsable et l'AI Act,
- des expertises en innovations scientifiques et techniques de pointe des partenaires de Dihnamic.

Au travers de **huit préconisations** détaillées dans ce document, la charte met en avant les axes essentiels **pour une collaboration « IA digne de confiance - entreprise »** respectant le droit et le bien-être des collaboratrices et collaborateurs dans les domaines public et privé.

Pour chacune d'entre elles, **des services concrets** sont proposés par les partenaires de Dihnamic aux expertises complémentaires - parmi lesquels l'acculturation, la formation, le conseil, le développement de prototype, la recherche de financement – afin de répondre aux besoins de chaque entreprise ou autorité publique de manière adaptée. Dihnamic se propose, au travers de cette charte, de combler le fossé entre la nouvelle réglementation européenne et les réalités logistiques, financières et industrielles des entreprises manufacturières au travers de la création d'**une vision commune en IA digne de confiance fédérant tout l'écosystème d'innovation**.

**Le centre Inria de l'université de Bordeaux
Manager IA digne de confiance, Dihnamic
pour les partenaires Dihnamic**
Avril 2025

² Nous pouvons citer 3 actions régionales : (i) la chaire industrielle sur l'IA digne de confiance fondée en 2023 par KEDGE, la Fondation Université de Bordeaux et l'école d'ingénieurs Enseirb Matmeca -Bordeaux INP, (ii) la charte Numérique responsable et les travaux de recherche de l'Institut du numérique responsable à la Rochelle, ainsi que (iii) les chaires 3IA portées par ANITI et l'Université de Toulouse pour une IA acceptable, une IA certifiable et une IA collaborative. Le lecteur pourra se référer à la section "Ressources supplémentaires" en fin de document pour plus d'informations

Hub d'experts Dihnamic

Un grand merci au groupe d'experts de haut niveau sur l'IA, présenté en suivant, pour sa contribution aux discussions, à l'analyse des recommandations et ses conseils sur la Charte et ses préconisations.

Catherine Tessier

Directrice de recherche, ONERA
Référente intégrité scientifique et
éthique de la recherche
Membre du Comité national pilote
d'éthique du numérique (2019-2024)

Claude Kirchner

Directeur de recherche émérite, INRIA
Président du Comité Consultatif
National d'Éthique du Numérique

Frédéric Alexandre

Directeur de recherche, INRIA
Responsable Equipe Mnemosyne
Chercheur en IA, modélisation
cognitive et neurosciences
computationnelles

Laurent Simon

Professeur des Universités à l'école
d'ingénieurs ENSEIRB-MATMECA de
Bordeaux INP et chercheur LaBRI à
l'Université de Bordeaux
Conseiller régional, CESER de Nouvelle-
Aquitaine
CHAIRE industrielle IA digne de confiance

Michèle Barbier

Coordinatrice de projet européen,
INRIA (jumeau numérique)
Experte en éthique pour la Commission
Européenne et les projets financés par
l'UE
Membre du conseil consultatif externe
de l'Ocean Entreprise Initiative

Nicolas Roussel

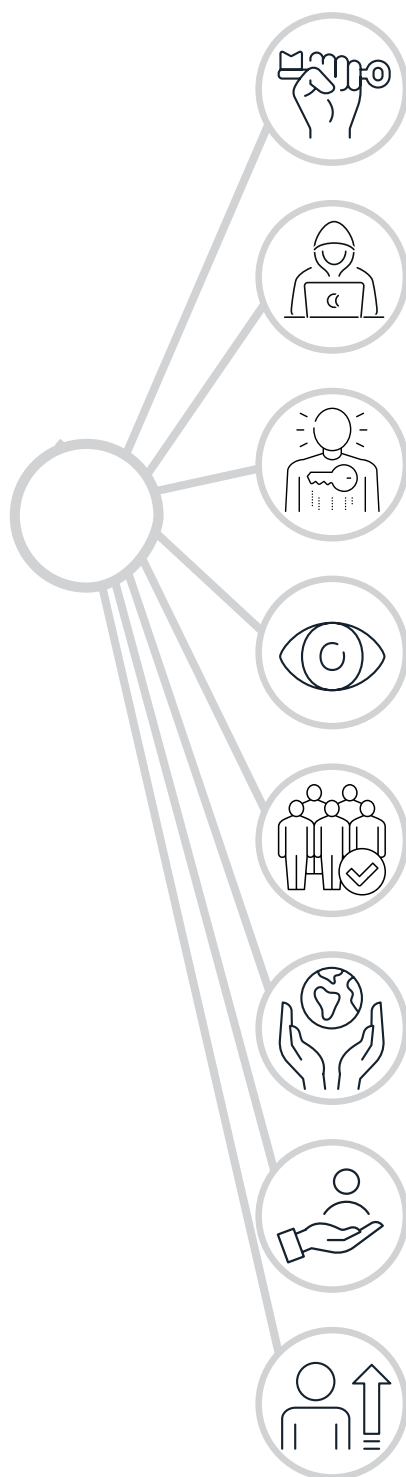
Directeur de recherche, INRIA
Chercheur en Interaction Humain-
Machine
Directeur du Centre Inria de l'université
de Bordeaux

Pierre Etienne Legoux

Responsable de l'IA, THALES AVIONICS

Charte Dihnamic

Intelligence Artificielle de Confiance & Éthique pour le Numérique



Action humaine et contrôle humain

Comprend les droits fondamentaux, l'autonomie humaine

Robustesse technique et sécurité

Comprend la résilience aux attaques et la sécurité, les plans de secours et la sécurité générale, la précision, la fiabilité et la reproductibilité

Respect de la vie privée et gouvernance des données

Comprend la qualité et l'intégrité des données, l'accès aux données

Transparence

Comprend la traçabilité, l'explicabilité et la communication

Non-discrimination, équité et diversité

Comprend l'absence de biais injustes, l'accessibilité et la conception universelle, et la participation des parties prenantes

Bien-être sociétal et environnemental

Comprend la durabilité et le respect de l'environnement, la frugalité, l'impact social, la société et la démocratie

Responsabilité (sociale et sociétale)

Comprend l'auditabilité, la réduction au minimum des incidences négatives et la communication à leur sujet, les arbitrages et les recours

Dignité

Comprend l'absence de toute sorte de malveillance ou d'atteinte à l'égard des êtres humains, la protection de leur intégrité physique et mentale, leur sentiment d'identité personnel et culturel et la satisfaction de leurs besoins essentiels

En détail



Action humaine et contrôle humain : une autonomie de prise de décision à tout moment

Afin de garantir le respect des droits fondamentaux, l'humain doit pouvoir prendre la décision de ne pas faire appel au système d'IA (SIA) s'il juge que les conditions éthiques ou de sécurité ne sont pas remplies. La gouvernance des SIAs, la protection de l'autonomie ainsi que de la prise de décision, le développement de conceptions centrées sur l'humain avec supervision et contrôle de l'IA, sont nécessaires.

Concrètement, il convient de :

- établir des protocoles pour qualifier le contexte de l'IA, le niveau de délégation et le rôle de l'humain (identifier les points de contrôle humain). En garantir le suivi ;
- mettre en place un processus de détection, d'alerte et de correction des défaillances (non-respect des droits fondamentaux, situation à risque) ;
- définir un processus de reprise en main du SIA en cas de défaillance ;
- définir les portes d'entrée d'urgence dans le SIA pour corriger, surveiller, émettre des rapports vers les utilisateurs.



Robustesse technique et sécurité

La fiabilité et la reproductibilité des décisions, face aux variations de contexte et aux perturbations, permettent de développer la robustesse et la résilience face aux attaques (problématiques liées à la cybersécurité, à l'imprévisibilité et à la volatilité de l'environnement réel où les SIA sont intégrés).

Concrètement, il convient de :

- sécuriser l'apprentissage : identifier les vulnérabilités du système IA (cas d'application et au-delà) et établir une check-list ;
- instaurer des protocoles de tests et de contrôle du comportement pendant toute la durée de vie du système pour s'assurer que le risque de dérive est impossible ;
- mettre en place des standards ou métriques d'évaluation sur la robustesse et définition d'un degré de tolérance à l'erreur ;
- mettre en place des seuils de performance permettant d'informer si le SIA a atteint ses limites de fonctionnement et prendre la main lorsque les données sont compromises ;
- maintenir la stabilité du système en veillant à la fiabilité des résultats et protection physique contre les imprévus ;
- surveiller et gérer l'ensemble des risques et anticipation des modifications de l'environnement (procédures adaptées communiquées en amont).



Respect de la vie privée et gouvernance des données

La protection des données privées, leur qualité et leur intégrité, ainsi que leur gouvernance des données en conformité avec le "*Data Governance Act*"³ de l'Union Européenne, garantissent le principe de prévention de toute atteinte à la vie privée des personnes.

Concrètement, il convient de :

- s'assurer de la conformité légale et réglementaire du traitement des données à caractère personnel en lien avec le règlement général de protection des données, RGPD⁴ (contrôles légalement obligatoires) et au Data Governance Act / CNIL : anonymiser, effectuer une veille régulière pour préserver, dans le temps, le caractère anonyme des données produites ;
- faire apparaître de manière accessible, claire et intelligible aux personnes produisant les données qui sont traitées par le SIA les conditions légales et réglementaires de collecte, stockage, traitement et partage. Obtenir leurs consentements ;
- mise en place d'une gouvernance des données comprenant des méthodologies de traçabilité du cycle de vie des données depuis la récolte/génération à leur utilisation, méthodologie de vérifications, procédures et opérations impactant ou utilisant les données.



Transparence : la traçabilité et l'explicabilité pour comprendre, expliquer et justifier un résultat

Pouvoir justifier, comprendre, interpréter les modèles et méthodes sous-jacents au SIA, permet à l'humain d'expliquer les résultats obtenus par le système, son comportement, et favoriser son adoption. Tracer les données utilisées permet également d'expliquer le comportement des SIAs, les résultats et les rendre transparents et plus compréhensifs.

Concrètement, il convient de :

- mettre en place et tenir à jour un référentiel dans lequel seront stockées toutes les informations liées à l'IA : expliquer la conception du modèle, sa performance, les résultats obtenus, ainsi que les limites techniques et choix techniques ;
- développer la documentation relative et préparer des schémas de vulgarisation, utiliser des outils graphiques qui permettent de visualiser les critères prépondérants ;
- mise en place de standards internes / de règles avec révision régulière pour anticiper tout risque ;
- justifier le fonctionnement du SIA : comparer les opérations de l'IA au raisonnement d'un humain dans le même cas d'usage ;
- évaluer et documenter le compromis performance vs transparence lors de la mise en place d'un SIA ;
- veiller à générer des explications appropriées des SIAs, adaptées au(x) besoin(s) exprimé(s) du ou des public(s) cible(s) des processus d'explicabilité et évaluer régulièrement leur acceptabilité et intelligibilité.

3 <https://www.cnil.fr/fr/strategie-europeenne-pour-la-donnee-la-cnil-et-ses-homologues-se-prononcent-sur-le-data-governance> (Dernier accès : 01/04/2025)

4 <https://www.economie.gouv.fr/entreprises/reglement-general-protection-donnees-rgpd> (Dernier accès : 01/04/2025)



Non-discrimination, équité et diversité : garantir un SIA éthique pour tous

Selon les contraintes métiers, technologiques et le niveau de risque associé au SIA, il convient de limiter la présence de biais potentiels pouvant nuire à une population donnée. Dans le cas échéant, il faut les documenter et informer les parties prenantes impactées. En outre, afin de garantir un accès équitable et la prévention de toute atteinte à l'humain, il convient aussi d'assurer l'accessibilité du SIA, sa conception universelle et la participation de toutes les parties prenantes potentiellement impliquées et impactées par le SIA tout au long du cycle de développement logiciel depuis les premières phases d'idéation – analyse des besoins – à la phase de déploiement.

Concrètement, il convient de :

- mettre en place des méthodologies de détection et de limitation des biais statistiques dans les jeux de données au travers de la mise en place de méthodologie de vérification/validation des modifications liées aux biais et des décisions ;
- sensibiliser ses équipes aux biais cognitifs qui interviennent lors de l'idéation, la conception et le test du SIA ;
- choisir et justifier le type de transformation équitable à adopter lorsque le SIA amène une prise de décision sur une population donnée : collaborer avec des spécialistes des sciences humaines et sociales et/ou des statisticiens pour élaborer un système éthique ;
- pratiquer et publier des auto-évaluations régulières de la conformité des travaux des data-scientists à la régulation, aux valeurs de l'entreprise et selon les contraintes techniques.



Bien-être sociétal et environnemental : deux missions de tout SIA responsable

L'IA durable et respectueuse de l'environnement doit veiller à ne pas nuire à la société et la démocratie, et inclure la frugalité énergétique pour minimiser l'incidence environnementale. Les applications d'IA doivent aider à construire un futur désirable par l'entreprise, ses consommateurs et les citoyens.

Concrètement, il convient de :

- identifier les risques du SIA sur les fonctions cognitives et la santé mentale des utilisateurs et les éliminer ou les réduire au minimum ;
- identifier et définir des valeurs socio-économiques à respecter dans le cadre de son développement et de son utilisation ;
- identifier des métriques d'évaluation/méthodologies de prise en compte des impacts (choix des fournisseurs, prestataires et sous-traitants impliqués dans le cycle de vie du SIA, consommation carbone, mise en place d'une politique d'archivage, d'expiration et de suppression des données) ;
- les appliquer, selon les contraintes techniques du domaine et de l'entreprise, au niveau des algorithmes et généraliser à tous les cas d'application ;
- sensibiliser les équipes et les utilisateurs à l'impact durable : environnement, santé et socio-économie ;
- appliquer les principes de bioéthique si le SIA a une vocation santé.



Responsabilité : identifier les responsabilités sociale et sociétale à tout moment

La responsabilité d'un acteur de la chaîne de production sur un SIA entraîne la nécessité pour celui-ci d'en assumer les conséquences. L'audibilité, la documentation et la communication des incidences négatives, celle des arbitrages et des recours permettent de garantir l'autonomie et la responsabilité à l'égard des systèmes d'IA et de leurs résultats.

Concrètement, il convient de :

- mettre en place des protocoles pour qualifier le contexte de l'utilisation du SIA, niveau de délégation et le rôle de l'humain et définir une procédure d'identification des responsabilités dans la chaîne de conception-mise en œuvre d'un SIA ;
- déployer les protocoles et les intégrer dans tous les projets IA : nommer un « model owner » (ou personne en charge du modèle) qui sera « responsable » du modèle/système en production ;
- définir les motifs susceptibles d'engager la responsabilité de chaque acteur de la chaîne : du concepteur / du développeur / du fournisseur / de l'hébergeur / des autres sous-traitants / de l'utilisateur final ;
- mettre en place des formations d'initiation et mises à jour aux technologies du numérique en lien avec les SIAs de la cybersécurité, l'explicabilité, la frugalité, l'interaction humain-machine.



Dignité : respect de la dignité Humaine

L'absence de toute sorte de malveillance ou d'atteinte à l'égard des êtres humains, la protection de leur intégrité physique et mentale, leur sentiment d'identité personnelle et culturelle et la satisfaction de leurs besoins essentiels sont indispensables au respect de la dignité humaine. De manière plus générale, un système d'IA peut se qualifier d'éthique s'il est en mesure de préserver, tout au long de son cycle de vie, les droits humains fondamentaux : droits à la dignité, à l'intégrité mentale et physique, à la liberté, à l'autonomie, à l'équité de traitement, à l'intimité et à la vie privée.

Concrètement, il convient de :

- qualifier et évaluer le SIA au regard des droits fondamentaux ;
- éduquer les consommateurs/employés sur les capacités et les limites des systèmes d'IA qu'ils utilisent. Recueillir leurs consentements à une automatisation et à sa finalité ;
- réaliser un guide d'évaluation du SIA pour l'utilisateur final ;
- mettre en place un message d'information à destination de l'utilisateur final qui soit visible, intelligible et paramétrable.

Glossaire

Intelligence Artificielle

Dans le contexte de cette charte, l'Intelligence Artificielle, est définie comme tout outil utilisé par une machine capable de "reproduire des comportements liés aux humains, tels que le raisonnement, la planification et la créativité"⁵. Les champs d'action de l'IA sont vastes, car s'étendent à de nombreux aspects du quotidien, innervent tous les domaines de la société, et représentent un enjeu technologique qui concerne l'économie, la recherche, la formation et nécessite par extension un ensemble de mesures et d'approches pour être dignes de confiance.

Numérique Responsable

Le Numérique Responsable est une démarche d'amélioration continue qui concerne l'empreinte écologique et sociale du numérique. Cela recouvre le Green IT pour réduire l'empreinte environnementale à l'échelle des Directions des systèmes informatiques (DSI) et l'IT for green qui met le numérique au service du développement durable et la conception responsable des services numériques.

La mise en place d'une démarche Numérique Responsable a donc en résumé pour objectifs de :

1. réduire l'impact environnemental ;
2. améliorer l'impact social ;
3. devenir un levier d'économie ;
4. être un levier d'innovation ;
5. favoriser l'engagement des collaborateurs et entreprises au travers des démarches Responsabilité Sociétale des Entreprises RSE.

Comment citer cette charte ?

EDIH Dihnamic (Avril 2025) « Pour une industrie de production numérisée, résiliente et éthique en Nouvelle-Aquitaine, CHARTE IA DIGNE DE CONFIANCE 2025 ».

La charte Dihnamic est publiée sous la licence :
CC BY-NC-SA 4.0 - Attribution-NonCommercial-ShareAlike 4.0 International.

⁵ Ministère chargé de l'enseignement supérieur et de la recherche (Mai 2024), "Intelligence artificielle (IA) : de quoi parle-t-on ?" [site Web]. Consulté le 03/04/2025 : <https://www.enseignementsup-recherche.gouv.fr/fr/intelligence-artificielle-de-quoi-parle-t-91190>

Pour aller plus loin

Pour aller plus loin, ci-dessous des ressources supplémentaires – non exhaustives – ayant aidé au travail de réflexion et de création de cette charte (dernière date d'accès à l'ensemble de ces liens : le 01/04/2025).

Au niveau régional

- La chaire industrielle sur l'IA digne de confiance fondée en 2023 par KEDGE BUSINESS SCHOOL, la Fondation Université de Bordeaux et l'école d'ingénieurs ENSEIRB Matmeca de Bordeaux INP : <https://www.fondation.univ-bordeaux.fr/projet/chaire-ia-digne-de-confiance>
- Les travaux et la charte Numérique responsable de l'Institut de Recherche en Numérique Responsable à La Rochelle : <https://institutnr.org/charte-numerique-responsable>
- Les chaires 3IA portées par ANITI et l'Université de Toulouse. 3 programmes de recherche dédiés à une IA acceptable, une IA certifiable et une IA collaborative : <https://aniti.univ-toulouse.fr/chaieres-3ia/>
- Les chaires IA Cluster ANITI renouvelées en 2024 : <https://aniti.univ-toulouse.fr/les-chaieres-ia-cluster-aniti/>

Au niveau national

- La circulaire n° 6425-SG du 21 novembre 2023 relative à l'engagement pour la transformation écologique de l'État : <https://www.legifrance.gouv.fr/download/pdf/circ?id=45511>
- Le site MiNumEco de la mission interministérielle numérique écoresponsable : <https://ecoresponsable.numerique.gouv.fr/a-propos/>
- Le guide pratique pour des IA éthiques de Numeum (2021) : <https://ai-ethical.com/wp-content/uploads/2021/09/2021-SN-Guide-Me%cc%81thodo-IA-Ethiques-version-imprime%cc%81e.pdf>
- Le livre blancs du Hub France IA (2023) : https://www.hub-franceia.fr/wp-content/uploads/2023/05/Livre_Blanc_IA_Ethique.pdf
- Le livre blanc du programme Confiance.ai : <https://www.confiance.ai/wp-content/uploads/2023/09/LivreBlanc-Confiance.ai-Octobre2022-1.pdf>

Au niveau européen et international

- Les lignes directrices en matière d'éthiques publiées par le High-Level Expert Group de la Commission Européenne en 2018 : <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- Les préconisations en législations européennes sur l'IA et l'AI Act : <https://digital-strategy.ec.europa.eu/fr/policies/regulatory-framework-ai>
- Les préconisations de l'UNESCO sur l'éthique en IA : <https://www.unesco.org/fr/artificial-intelligence/recommendation-ethics>



www.dihynamic.eu



DIHNAMIC



@dihynamic



contact Dihynamic
contact@dihynamic.eu

contact Charte
Ikram Chraibi Kaadoud, Inria
ikram.chraibi-kaadoud@inria.fr

13 partenaires à votre service



Partenaires associés :

Agri Sud-Ouest Innovation | Cosmetic Valley |
École Nationale Supérieure d'Arts et Métiers (campus Bordeaux-Talence) |
Pôle Européen de la Céramique | Université de Bordeaux

Cofinancé par :



Cofinancé par
l'Union européenne



Projet cofinancé par la
Région Nouvelle-Aquitaine

Project n° 101083886 Dihynamic

Avertissement : Le projet Dihynamic est cofinancé par l'Union Européenne au titre du programme Digital Europe et par la Région Nouvelle-Aquitaine. Les points de vue et les opinions exprimés ici n'engagent que les auteurs et ne reflètent pas nécessairement ceux de l'Union européenne ou de l'Agence Exécutive européenne de la Recherche. Ni l'Union européenne ni l'agence ne peuvent être tenues pour responsables de ces opinions.