

 [VU, LU, ENTENDU](#)

Données synthétiques : baguette magique ou mirage ? – Tribune de Mathilde Guinaudeau, Ipsos bva

6 Jan. 2026

FUTUR DU MARKET RESEARCH | IA ET MARKET
RESEARCH | IPSOS | TRIBUNES / POINTS DE VUE



Partager

Facebook

Twitter

LinkedIn

Après les tribunes de [Luc Balleroy](#) (OpinionWay) et [Laurent Florès](#) sur les interviews synthétiques — et plus largement l'usage de l'IA — dans le domaine des études marketing et opinion, le débat s'enrichit d'une troisième perspective. Mathilde Guinaudeau, Head of Data & AI Labs chez [Ipsos bva](#), pose une question essentielle : comment utiliser ces données synthétiques de manière rigoureuse et responsable, en dépassant les discours simplistes qui oscillent entre promesses miraculeuses et mises en garde excessives ? Sa réponse s'appuie sur une distinction fondamentale : derrière ce terme générique se cachent en réalité quatre catégories aux usages et exigences de validation très différents. Nous vous proposons de découvrir son analyse et ses recommandations.

LA TRIBUNE DE MATHILDE GUINAudeau, HEAD OF DATA & AI LABS CHEZ IPSOS BVA

Les données synthétiques révolutionnent les possibles tout en s'appuyant sur les fondamentaux

L'année démarre avec un défi aussi stimulant que stratégique pour notre secteur : **démystifier les idées reçues autour des données synthétiques**. À mesure que ces dernières gagnent en visibilité, nous observons qu'elles cristallisent des discours souvent simplistes et polarisés, entre promesses quasi

miraculeuses et mises en garde excessives. Rappelons d'abord l'essentiel, les données synthétiques ne signent pas la fin des sondages dits « classiques », ni à l'inverse, ne conduisent à des interprétations artificielles ou biaisées lorsqu'elles sont correctement produites et utilisées.

Ma conviction est claire : **elles ouvrent au contraire de nouvelles opportunités et perspectives**, en apportant flexibilité, rapidité et profondeur d'analyse. Mais à une condition non négociable : maintenir l'humain, l'expertise métier et la rigueur scientifique au cœur des décisions et des processus de validation.

Des promesses fortes qui ne naissent jamais ex nihilo

Nous constatons aujourd'hui un discours très volontariste autour de la capacité à produire des données synthétiques rapidement, parfois en très grand volume. Certains nouveaux entrants affirment pouvoir générer en quelques heures des centaines, voire des milliers de persona. Cette promesse appelle une question fondamentale : **sur quelles bases et avec quelles garanties** ? Ces 2 interrogations majeures doivent systématiquement être posées.

La première concerne la validation : quels critères sont utilisés ? Quels niveaux de fiabilité sont atteints ? Avec quels seuils d'erreur acceptables ?

La seconde porte sur l'origine des données : sur quels corpus réels les modèles sont-ils entraînés ? Qui détient ces données ? Comment ont-elles été collectées et qualifiées ? Car une réalité s'impose : **les données synthétiques ne s'auto-génèrent pas**. Elles sont toujours issues, directement ou indirectement, de données collectées auprès de personnes bien réelles.



Ecouter aussi > [Le podcast de MRNews – Données synthétiques : quelle place dans les études marketing ? Avec Stéphane Marcel, Président de Syntec Etudes](#)

Derrière le terme « données synthétiques », une réalité plurielle

Autre écueil fréquent : parler des données synthétiques comme d'un ensemble homogène. En pratique, nous distinguons **au moins quatre grandes catégories** : le Data Boosting, la Data Imputation, les Persona Bots et les Digital Twins, chacune correspondant à des usages spécifiques et à des exigences de validation différentes.

1. Le Data Boosting

Le Data Boosting consiste à **augmenter la taille de sous-échantillons** au sein d'un jeu de données existant. Il permet, par exemple, de répondre à des questions comme « J'aimerais avoir plus d'échantillons de la marque X ou de la tranche d'âge Y » **pour pouvoir produire des résultats statistiquement significatifs**. Cette approche permet de générer des répondants simulés, uniques et distincts, à partir des réponses déjà collectées. En revanche, **elle ne permet pas d'introduire de nouvelles questions**.

Son efficacité repose sur plusieurs prérequis : un échantillon principal suffisamment large (au moins 500 interviews, idéalement 1 000), un nombre minimal de répondants au sein du sous-échantillon à booster (minimum 20 répondants) et un facteur de multiplication contrôlé (maximum x3). Toutefois, la fiabilité se dégrade lorsque le data boosting est appliqué à des questionnaire plus complexes, notamment avec des questions filtrées, comme par exemple les funnels d'achat.



2. La Data Imputation

La Data Imputation vise à **compléter des données manquantes** ou à simuler des réponses non collectées. Elle s'applique notamment lorsque certaines questions ont été retirées ou ajoutées à un baromètre : « J'ai cessé de poser une question dans mon baromètre et je souhaite connaître la réponse potentielle de ma nouvelle vague » ou « J'ai introduit une nouvelle question et je veux savoir ce que mon échantillon historique aurait pu répondre ». Elle offre également un bénéfice opérationnel majeur : **la réduction significative de la longueur des questionnaires** (jusqu'à -40%), limitant la fatigue des répondants et améliorant, par ricochet, la qualité des réponses.

3. Les Persona Bots

Les Persona Bots sont des **assistants numériques incarnant des segments spécifiques de consommateurs**. Ils agrègent les traits, comportements ou attributs communs observés au sein d'un groupe d'individus, qu'il s'agisse d'un segment de clients d'une marque par exemple ou d'un archétype, comme la Génération Z ou les citoyens soucieux de l'environnement. Contrairement à certaines idées reçues, ils ne reposent pas sur un simple prompt : ils sont entraînés sur des données réelles, quantitatives et qualitatives, constituant une base de connaissances qui leur permet de refléter fidèlement les logiques observées sur le terrain. Ils permettent d'interagir de manière fluide avec un segment donné, de poser de nouvelles questions et de faciliter l'appropriation des cibles au sein des organisations. Leur valeur est particulièrement forte pour la génération d'idées, l'exploration de scénarios et le développement de concepts produits ou de plateformes de communication.

En revanche, ils ne remplacent en aucun cas une étude quantitative. Leur fiabilité est strictement limitée au périmètre des données sur lesquelles ils ont été entraînés. La validation repose sur des critères

mixtes, à la fois qualitatifs et quantitatifs, intégrant des scores de pertinence et d'adéquation des réponses au périmètre de l'analyse.

4. Les Digital Twins

Les Digital Twins représentent l'approche la plus avancée : **une réplique synthétique d'un individu réel**. Ils permettent de poser de nouvelles questions sans nécessairement retourner immédiatement sur le terrain, ou d'interroger plus fréquemment des populations rares ou difficiles d'accès (B2B, clients du luxe etc...), tout en s'affranchissant de la fatigue des répondants. Là où un individu réel atteint rapidement un seuil de saturation, un Digital Twin peut simuler une infinité de scénarios sans jamais s'épuiser. En reproduisant les comportements, les caractéristiques et les réponses spécifiques d'un individu dans différents contextes, les Digital Twins rendent notamment possible la réalisation de nouveaux tests quantitatifs.

Cette promesse et sa fiabilité reposent toutefois sur un prérequis non négociable : un **échantillon initial extrêmement robuste** et une phase de **validation systématique**, comparant les réponses synthétiques à celles de répondants réels pour garantir qu'elles ne sont pas de simples hallucinations statistiques, mais bien des prédictions fidèles.

C'est dans cette logique qu'Ipsos a engagé une collaboration avec le *Politics and Social Change Lab* (PASCL) de l'Université de Stanford, afin de développer des méthodologies avancées et des cadres de validation propriétaires, intégrant à la fois les dimensions méthodologiques, technologiques et éthiques.



Pourquoi L'humain reste incontournable

Ces quatre catégories partagent un point commun fondamental : **leur validation repose toujours sur une double exigence, académique et professionnelle**. Elles mobilisent à la fois la recherche universitaire et l'expertise des spécialistes en intelligence artificielle, data science, statistiques et mathématiques.

Plus profondément encore, **leur origine reste humaine**. Les données synthétiques ne sont jamais « hors-sol » : elles prolongent des comportements, des opinions et des expériences vécues. Sans cet ancrage dans le réel, leur fiabilité s'effondre.

C'est dans cette logique qu'Ipsos s'appuie notamment sur le Knowledge Panel, l'un des plus grands panels probabilistes en ligne d'Europe. Il permet un contrôle rigoureux de la qualité, de la représentativité et du renouvellement des échantillons, incluant des publics éloignés du numérique. Le fait d'être le seul institut d'études propriétaire de ses panels permet à Ipsos d'en assurer la maîtrise complète. Ce socle est indispensable pour produire des données synthétiques robustes, sourcées et responsables.

Enfin, un dernier enjeu s'impose à nous collectivement : **l'évaluation des données synthétiques elles-mêmes**. Les référentiels traditionnels ne suffisent plus. Il nous appartient de définir de nouveaux standards, de clarifier les marges d'erreur acceptables, les risques associés et la place de ces données dans la prise de décision.

Car, in fine, **la responsabilité de l'interprétation et de la décision reste fondamentalement humaine**.



Lire aussi, sur le même sujet > La tribune de Luc Balleroy (DG d'OpinionWay) : « [Ne laissons pas le mirage des interviews synthétiques brouiller la révolution IA](#) »

et celle de Laurent Florès (Université Paris Assas & L'Atelier IA) : « Études marketing et IA : la cinquième génération ne sera pas une évolution, mais une révolution »

POUR ACTION

- Echanger avec l'interviewé(e) : [@ Mathilde Guinaudeau](#)
- Echanger sur LinkedIn au sujet de cette tribune

[En voir + sur Ipsos](#)

Vous avez apprécié cet article ? N'hésitez pas à le partager !

[Facebook](#)

[Twitter](#)

[LinkedIn](#)

CET ARTICLE VOUS A INTÉRESSÉ ?

Tenez-vous régulièrement informé de notre actualité et de nos prochains articles en vous inscrivant à notre newsletter.



[S'ABONNER A LA NEWSLETTER](#)

Ces thèmes et articles pourraient également vous intéresser...

FUTUR DU MARKET RESEARCH | IA ET MARKET RESEARCH | IPSOS | TRIBUNES / POINTS DE VUE

« Même un très bon nom de marque a besoin de temps pour s'imposer » – Interview de Nadège Depeux (Labbrand)

Données synthétiques : baguette magique ou mirage ? – Tribune de Mathilde Guinaudeau, Ipsos bva

En tête à tête avec Didier Truchot, Président et Fondateur d'Ipsos – Le 'replay' de la grande interview du Printemps des études

RETRouvez tous nos articles vu-lu-entendu

Pourquoi MRNews ?

Qui sont les lecteurs de
MRNews ?

Les articles les plus lus en 2024

Nous contacter

Rejoindre le Club MRNews

Mentions légales

Suivre MRNews sur LinkedIn



Pour rester informé(e) de notre
actualité et recevoir notre
newsletter mensuelle :

Email

S'ABONNER

Confidentialité - Conditions

Conception Authentic Mood – © Atelier Eclectik – Tous droits réservés

