



Thématique :

IAFLUENCE

DEPP



CONTACT POUR LE CHALLENGE

Nom, Prénom : ROCHER, Thierry

Fonction : Adjoint au sous-directeur des évaluations et de la performance scolaire, DEPP

Numéro de téléphone : 060744044

RESUME DU CHALLENGE

Le challenge consiste à **élaborer un outil de correction automatique de la fluence** (rapidité et précision de lecture à l'oral). Le lauréat bénéficie de l'accès à un jeu de données d'apprentissage : les voix de 25 000 élèves enregistrées et codées par la DEPP, pour construire un algorithme IA. Cet outil, qui n'existe pas aujourd'hui, permettrait aux enseignants d'évaluer efficacement le niveau de fluence de lecture de leurs élèves. La solution pourra utilement s'intégrer dans des logiciels éducatifs, au-delà du challenge.

Sommaire

1. CONTEXTE ET OBJECTIFS DU CHALLENGE	3
2. PRESENTATION DU SPONSOR ET EXPRESSION DU BESOIN.....	3
3. ÉTAT DE L'ART ET CHALLENGE A RESOUDRE	3
4. RESULTATS ATTENDUS.....	3
5. ENGAGEMENT DU SPONSOR, ÉQUIPE MOBILISEE ET MODALITES OPERATIONNELLES	3
6. RESSOURCES ET/OU JEUX DE DONNEES PROPOSES PAR LE SPONSOR.....	4
7. DEROULEMENT DU CHALLENGE - METHODOLOGIE.....	4
8. CRITERES DE SELECTION	4
9. RESTITUTION DU CHALLENGE	4
10. PERSPECTIVES ET RETOMBEES POSSIBLES DU CHALLENGE POUR LE LAUREAT	4

1. CONTEXTE ET OBJECTIFS DU CHALLENGE

Depuis 2015, les évaluations standardisées menées par la DEPP (service statistique du ministère de l'éducation nationale) ont opéré une transition vers le tout numérique. Dans ce cadre, la DEPP a conduit des évaluations à grande échelle grâce à des applications sur tablettes.

Ainsi, les voix de 25 000 élèves ont été enregistrées et corrigées (en partie) par la DEPP, lors d'une évaluation de la fluence : lecture à voix haute de mots et lecture à voix haute de textes.

Le challenge a pour but de s'appuyer sur ce corpus de données pour construire une solution permettant de positionner un élève automatiquement en matière de fluence (nombre de mots correctement lus en une minute), à partir de l'enregistrement de sa lecture orale.

2. PRESENTATION DU SPONSOR ET EXPRESSION DU BESOIN

La **Direction de l'évaluation, de la prospective et de la performance (DEPP)** est à la fois une direction du ministère en charge de l'Éducation nationale et le service statistique ministériel en charge des statistiques sur l'éducation : <https://www.education.gouv.fr/direction-de-l-evaluation-de-la-prospective-et-de-la-performance-depp-12389>

Concernant les **évaluations standardisées des élèves**, la DEPP conduit de nombreux programmes :

- évaluation de l'ensemble des élèves de CP, CE1, 6^{ème}, 2^{nde}, JDC
- enquêtes nationales sur de larges échantillons (CEDRE, Socle, Panels, Évaluation de politiques publiques)
- enquêtes internationales sur de larges échantillons (PISA, PIRLS, TIMSS, ICILS)

L'ensemble de ces programmes d'évaluations standardisées concernent au total environ 4 millions d'élèves chaque année. L'équipe chargée du développement de ces programmes est composée d'environ 50 personnes, dont 15 statisticiens/psychométriciens.

Depuis 2015, ces évaluations ont opéré une **transition vers le tout numérique**. Cette transformation pose de nouveaux défis mais fait naître de nouvelles opportunités, notamment en matière d'**automatisation de corrections**.

De ce point de vue, une dimension est particulièrement intéressante en matière d'automatisation : la **fluence de lecture**. Traditionnellement, la fluence de lecture est captée à travers le *One-minute test* : on demande à l'élève de lire à voix haute et le plus rapidement possible une liste de mots ou bien un texte. Le correcteur est chargé de compter le nombre de mots lus correctement en une minute, qui sert d'indicateur de fluence.

La fluence de lecture est une dimension essentielle à maîtriser pour que les élèves puissent suivre une scolarité de manière efficace. Or, à la rentrée 2020, pour la première fois, un test de fluence a été proposé à l'ensemble des 800 000 élèves de sixième. Ce test a montré que seule la moitié des élèves parvenait à atteindre le seuil de 120 mots minutes, requis en fin d'école primaire. Et 15 % d'entre eux n'ont pas atteint le seuil de 90 mots, seuil requis en fin de CE2 ! Ces élèves auront de grandes difficultés à poursuivre leur scolarité au cours du collège (ce taux est de 30 % en REP+)¹.

Afin de faciliter la passation de ces tests et donc d'améliorer leur acceptabilité et leur usage, une version automatisée de l'évaluation de la fluence est souhaitable.

¹ <https://www.education.gouv.fr/evaluations-de-debut-de-sixieme-2020-premiers-resultats-307125>

3. ÉTAT DE L'ART ET CHALLENGE A RESOUDRE

Dans le but d'évaluer cette dimension plus tôt dans la scolarité, et dans la perspective d'automatiser la mesure, la DEPP a enregistré les voix d'environ 25 000 élèves, lors d'une évaluation de la fluence de mots et de la fluence de texte. Un sous-ensemble de ces productions a été corrigé par une équipe d'enseignants recrutés par la DEPP.

Le challenge consiste donc à **élaborer un outil de correction automatique de la fluence** : l'élève lit une série de mots ou un texte, le système enregistre l'élève et « score » automatiquement son niveau de fluence, après traitement vocal. Les **voix enregistrées et codées de 25 000 élèves par la DEPP** peuvent servir de base à des données d'apprentissage pour un algorithme permettant une correction automatique.

Le challenge a pour but de s'appuyer sur ce corpus de données pour construire un outil numérique permettant de positionner un élève automatiquement en matière de fluence (nombre de mots correctement lus en une minute).

4. RESULTATS ATTENDUS

L'enjeu est de pouvoir construire un algorithme permettant un traitement automatique de ces enregistrements, sachant qu'ils ont été réalisés en situation quotidienne (micro de tablette, dans une classe de cours avec bruits environnants) et qu'ils concernent des voix d'enfants, dont on sait qu'elles sont moins facilement traitables informatiquement que les voix d'adultes.

Cet algorithme doit être intégré à un applicatif permettant son intégration dans des chaînes de production du sponsor et l'autonomie du sponsor vis-à-vis du traitement des données futures.

L'objectif final est de pouvoir appliquer le *scoring* automatique des voix en temps réel, par le biais de son intégration dans une application tablette ou dans un outil d'évaluation en ligne (la plateforme TAO, <https://www.taotesting.com/fr/>). A minima, une étude de faisabilité d'une telle solution synchrone intégrée est attendue.

Toute solution développée doit être accompagnée de la documentation.

5. ENGAGEMENT DU SPONSOR, ÉQUIPE MOBILISEE ET MODALITES OPERATIONNELLES

La DEPP investit beaucoup dans des solutions technologiques innovantes en matière d'évaluation (jusqu'à un engagement fort au niveau international à travers l'association FLIP+). Dans ce secteur, la correction automatisée de production d'élèves occupe une place importante et un secteur stratégique pour la DEPP en tant que service statistique.

L'équipe mobilisée pour ce projet sera composée de :

- Une cheffe de projets, en charge des programmes innovants
- Une data scientist
- Une professeure des écoles, spécialiste de l'évaluation numérique
- Un psychométricien

En outre, le service informatique de la DEPP assistera l'équipe sur les aspects liés aux données (voix).

Challenges Education – règlement challenge : IAFLUENCE

6. RESSOURCES ET/OU JEUX DE DONNEES PROPOSES PAR LE SPONSOR

Les données dont dispose la DEPP sont constituées des voix d'élèves de 25 000 élèves de CP ou de CE1 (échantillon représentatif) dans le cadre d'une évaluation de fluence de mots et de fluence de texte.

L'évaluation consiste à lire, pendant une minute, une liste de mots ou un texte. La liste de mots à lire est identique pour tous les élèves. En revanche, le texte à lire est différent pour les élèves de CP et les élèves de CE1.

Les élèves ont été évalués à plusieurs reprises. Le corpus total est composé de plus de 61 000 enregistrements (mots et textes) sous format mp3, correspondant à 9Go de données (compressées) :

- 25 000 enregistrements de la lecture d'une liste de mots (60 mots)
- 13 000 enregistrements de la lecture du texte "Dino" (niveau CP)
- 23 000 enregistrements de la lecture du texte "Géant" (niveau CE1)

Un sous-ensemble de ces enregistrements a été corrigé par des équipes d'enseignants à la DEPP. Les correcteurs ont effectué :

- 500 transcriptions des textes lus (250 CP, 250 CE1) : il s'agit de la transcription littérale de ce que l'élève a prononcé ;
- 7000 annotations des textes lus (3500 CP, 3500 CE1) à l'aide d'un outil de correction en ligne. L'outil permet aux codeurs de marquer les mots mal prononcés, d'indiquer le dernier mot lu ainsi que de renseigner le niveau (4 ou 5 catégories selon les dimensions) de prosodie et de juger sur la qualité audio.

Ces données sont la propriété de la DEPP et feront l'objet d'une convention de mise à disposition avant transmission au lauréat. Le travail pourra démarrer dès la mise à disposition de ces données.

7. DEROULEMENT DU CHALLENGE - METHODOLOGIE

T0 : Lancement de projet. Mise à disposition des enregistrements des voix et des corrections associées par le sponsor.

T0 + 2 semaines : Résultats du contrôle de qualité de données (voix, corrections), identification de besoins en termes de campagnes de correction*.

T0 + 1 mois : Présentation du plan d'action de traitement des voix et de solutions méthodologiques et techniques envisagées.

T0 + 2 mois : Résultats des premières calibrations d'algorithme et point de régulation*.

T0 + 3 mois : Présentation des résultats de l'algorithme et proposition de l'architecture de l'appliquatif ainsi que d'une étude de faisabilité de l'intégration du *scoring* automatique (en temps réel) aux plateformes/applications d'évaluations du sponsor.

T0 + 4 mois: *Mockup* de l'appliquatif.

T0 + 6 mois: *Minimum viable product* (MVP).

T0 + 7 mois: Résultats des tests et point de régulation*.

T0 + 9 mois: Livraison du POC - algorithme, applicatif, étude de faisabilité du *scoring* automatique en temps réel (service obligatoire), solution intégrée (service possible).

Des réunions de suivi auront lieu toutes les deux semaines entre le lauréat et la DEPP.

*En cas de besoin d'organiser de campagnes de corrections supplémentaires par le sponsor, le calendrier sera adapté pour prendre en compte les délais nécessaires (de 1 à 3 mois pour une campagne de correction, estimation pour 7000 enregistrements, 15 codeurs).

Challenges Education – règlement challenge : IAFLUENCE

8. CRITERES DE SELECTION

1- Pertinence de la proposition de réponse (pondération : 30)

- **Innovation** : caractère innovant des travaux réalisés/proposés par le lauréat (pondération : 10)
- **Cohérence et réalisme** de la proposition du calendrier des travaux (pondération : 10)
- **Prise en compte de la problématique** posée par le sponsor : la solution proposée par le lauréat répond au challenge du sponsor et à ses critères spécifiques (pondération : 10)

Critères de sélection spécifiques :

- Méthodologie proposée (Agile), description de l'organisation générale des prestations et processus.
- Utilisation de langages et de technologies : JavaScript, PHP, Java, Python, R, HTML, JSON, XML).
- Facilité d'accès aux solutions et qualité de la documentation dans l'objectif de futures évolutions et utilisations par le sponsor.
- L'utilisation des technologies *open source* est un plus.

2- Compétences et engagements (pondération : 30)

- **Niveau d'engagement du candidat** : identification des personnes clés (pondération : 10) (nombre et qualité, fourniture des CV) et quotité de disponibilité dédiée au challenge, mise à disposition d'un interlocuteur identifié et unique pour favoriser les interactions avec le sponsor

Deux types de profils sont attendus :

Profil data scientist

- ☐ Maîtrise des méthodes de deep learning et des techniques du traitement de voix
- ☐ Usage expert des environnements Big Data
 - ☐ Apache Spark est particulièrement attendu
- ☐ Usage de langages maîtrisés en interne par le sponsor
 - ☐ R, Python, SQL
- ☐ Qualités pédagogiques, capacité de communication avec les experts du champ étudié
 - ☐ spécifiquement, appétence pour le champ de la didactique et des sciences de l'éducation

Profil architect/développeur logiciel

- ☐ Connaissance experte en systèmes d'information et conception de logiciels
- ☐ Connaissance approfondie des technologies web
- ☐ Connaissance approfondie en développement d'applications
- ☐ Connaissance approfondie de méthodologies et technologies de bases de données
- ☐ Connaissance approfondie en sécurité de données
- ☐ Connaissance architectures Big Data
 - ☐ spécifiquement, la maîtrise experte de Hadoop et Spark est demandée
- ☐ Usage de langages maîtrisés en interne par le sponsor
 - ☐ R, Python, JavaScript sont particulièrement attendus
- ☐ Connaissance des standards d'interopérabilité employés dans le domaine de l'EdTech (QTI, LTI...)

- **Maturité de la réflexion** (pondération : 10) et des connaissances du candidat sur le monde de l'éducation et ses enjeux ;

Cadres règlementaires et juridiques (code de l'éducation, RGPD, handiaccessibilité...) et cadre de confiance.

Connaissance des programmes scolaires.

Challenges Education – règlement challenge : IAFLUENCE

- **Ressources en démonstration** (pondération : 10) : indication de réalisations numériques déjà produites (dans l'éducation ou autre, produits industriels ou prototypes).

3 - Prix. (pondération : 40)

9. RESTITUTION DU CHALLENGE

Les services obligatoires attendus sont les suivants :

- Développement d'un algorithme d'IA permettant le *scoring* automatique des voix d'élèves.
- Développement d'un applicatif de l'algorithme permettant son intégration dans des chaînes de production du sponsor.
- Étude de faisabilité de l'intégration du *scoring* automatique en temps réel dans une APK ou à une plateforme d'évaluation en ligne (TAO).

Les services possibles attendus sont les suivants :

- Intégration du *scoring* automatique en temps réel des voix dans une APK ou à une plateforme d'évaluation en ligne (TAO).

Tous les codes sources ainsi que la documentation associée seront demandés au lauréat à l'issue du challenge.

La propriété intellectuelle des solutions développées doit appartenir au sponsor afin de garantir l'autonomie du sponsor au regard de leur implémentation sur des données futures.

10. PERSPECTIVES ET RETOMBÉES POSSIBLES DU CHALLENGE POUR LE LAURÉAT

Le lauréat s'inscrit dans un axe parmi les plus innovants au croisement du numérique et de l'éducation, de la recherche et de l'action pédagogique.

Le lauréat aura accès à un corpus unique de voix (61 000 enregistrements des voix d'élèves de CP et de CE1 représentatifs de l'ensemble des régions de France). A ce corpus, s'ajoutent les 3000 enregistrements des voix des élèves de troisième récoltés dans le cadre d'une évaluation des compétences langagières et de la littérature. Le lauréat bénéficiera en outre de l'expertise de la DEPP en matière d'évaluation (didactique, psychométrie en particulier), sur laquelle il pourra capitaliser pour d'autres activités, dans un domaine en plein essor.

L'expérience sera valorisable au niveau national – à travers l'élaboration de solutions logicielles par exemple – ainsi qu'au niveau international, à travers les services étendus à des institutions partenaires de la DEPP à l'étranger (notamment via l'association FLIP+ : <https://flip-plus.org/>).

Cette expérience lui permettra de se positionner de façon optimale dans le secteur de l'application d'algorithme IA dans le domaine de l'évaluation, en particulier dans le traitement des voix d'enfants. L'outil développé pourra s'intégrer dans des solutions diverses de e-learning.

En outre, la DEPP a une gamme de publications très variées : notes, articles, dossiers, rapports. Par ailleurs, les équipes de la DEPP interviennent fréquemment dans des conférences scientifiques et dans des conférences réunissant des professionnels de l'évaluation et publient régulièrement dans des revues scientifiques. En outre, la DEPP est très active dans les réseaux internationaux sur l'évaluation.

Tous ces outils seront mobilisés pour valoriser les travaux réalisés avec le lauréat.

Challenges Education – règlement challenge : IAFLUENCE