

## ROC pour l'analyse automatique de copies d'examen

### Contexte

L'automatisation de l'évaluation de copies d'examen présente plusieurs avantages comme le gain de propriétés désirables sur les évaluations produites (déterminisme, vérifiabilité...), ou le gain en efficacité sur la production des évaluations (possibilité de donner des retours quasi-instantanés même sur de très grands groupes d'apprenant-e-s).

La plupart des outils d'évaluation automatiques existants font l'hypothèse forte que les apprenants utilisent des systèmes informatiques pour réaliser leurs rendus. Il est cependant courant que des universités préfèrent utiliser des moyens « papier » pour réaliser les épreuves importantes (contrôles terminaux, concours d'entrée...) et ceci pour différentes raisons (maîtrise de la fiabilité des conditions d'examens, limitation de la triche...). Des outils comme AMC permettent d'automatiser certaines tâches de l'évaluation de copies en passant par des supports « papier ». Ces outils sont malheureusement limités par les types de questions permises (variantes de QCM) et sur leurs possibilités de composition (pas pensés pour que des programmes génèrent des exercices et les corrigent, pas pensés pour faire des *pipelines* d'analyse de données). Ces limites nous ont poussé à lancer le développement d'un nouvel ensemble de logiciels appelé Compasse pour générer et évaluer automatiquement des copies d'examen.

### Objectifs du projet

Un stage réalisé pendant l'été 2023 a permis de réaliser une preuve de concept du composant d'analyse automatique de copies d'examen de Compasse. Ce code est implémenté en Rust et est fortement basé sur OpenCV. Le but de ce projet est de développer une bibliothèque en Rust pour la partie « reconnaissance optique de caractères » afin qu'elle soit utilisée par le composant d'analyse automatique de copies. Cette bibliothèque devra au minimum permettre de *parser* les zones de formulaires d'identification des apprenant-e-s (numéro étudiant, prénom, nom, groupe...). Idéalement, elle pourra aussi *parser* les zones de texte libres que peuvent compléter les apprenant-e-s, que ces zones comportent une ou plusieurs lignes.

Les objectifs principaux de ce projet sont les suivants :

- Évaluer la pertinence de Tesseract comme moteur d'OCR pour Compasse.
- Développer une bibliothèque Rust qui réalise les différentes tâches d'OCR nécessaires à Compasse.
- Améliorer la qualité du composant d'évaluation automatique de Compasse.
  - Automatiser les tests de fonctionnement du composant.
  - Automatiser le lancement de ces tests via GitLab CI.