

Banque MP Inter-ENS - Session 2024

Rapport sur l'épreuve de TIPE d'informatique

ENS (Paris) - ENS de Lyon - ENS Paris-Saclay – ENS de Rennes

Coefficients (en pourcentage du total des points de chaque concours) :

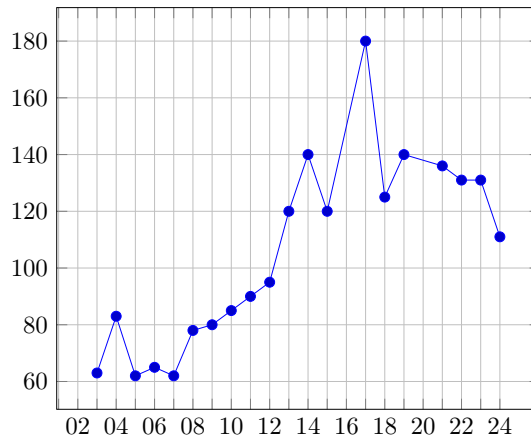
	ENS Paris-Saclay	ENS Lyon	ENS (Paris)	ENS Rennes
Concours MP :	3.8%	5.4%	7.4%	3.8%
Concours Info :	7.9%	5.6%	3.3%	—
Concours MPI :	7.9%	5.6%	3.3%	13.9%

Membres du jury : N. BERTRAND, P. SENELLART, É.-J. SIMS, N. SZNAJDER, É. THIERRY

1 Éléments de statistiques

Le jury a évalué cette année 111 candidates et candidats présentant un dossier TIPE d'informatique. Le nombre de candidates – 7 cette année – reste anormalement faible au regard de la proportion des chercheuses en informatique (environ 20%) qui est elle-même anormalement basse. Le jury encourage vivement les enseignantes et enseignants à diriger les candidates vers des sujets d'informatique. Cette année, les femmes étaient majoritaires au sein du jury et ce dernier a été attentif aux biais de genre dans ses appréciations.

Figure 1 – Nombre de candidats en TIPE d'informatique, par année.



Comme les années précédentes, le jury a apprécié la qualité d'un nombre important de TIPE, qu'ils s'inscrivent ou non dans le thème de l'année. Les candidates et candidats ont été évalué-es en premier lieu sur leur

maîtrise du sujet choisi, mais aussi sur leur connaissance des concepts généraux d'informatique et sur la qualité scientifique de leur travail. Le jury tient à souligner que le niveau des candidates et candidats auditionnés s'est significativement amélioré depuis une dizaine d'années. Une part importante des oraux étaient de niveau très satisfaisant et plusieurs de niveau vraiment excellent. Le jury a particulièrement apprécié la créativité et l'enthousiasme de certaines candidates et certains candidats, ainsi que les risques pris en explorant des thématiques nouvelles en lien avec les évolutions récentes de l'informatique. Toute la palette des notes a été utilisée. En conséquence, une note de 10 ou moins ne signifie pas nécessairement un oral de mauvaise qualité, mais uniquement un oral moins bon que la moyenne des oraux cette année.

2 Choix du sujet

Le choix du sujet est particulièrement important et doit permettre à la personne candidate de mettre en valeur sa créativité, la rigueur de son approche, son esprit critique et sa capacité à s'impliquer dans un projet qui la motive. L'originalité et la prise de risque sont souvent valorisées par le jury, surtout quand elles traduisent une implication particulière des candidates et candidats.

Le sujet doit mettre en avant un questionnement informatique, et donner lieu à une formalisation explicite du problème traité. Il doit par exemple permettre d'approfondir des notions de base de l'informatique, ou des questions d'algorithmique, des choix de structure de données, un domaine applicatif, etc. Chaque fois que le sujet s'y prête, un travail expérimental et une réalisation logicielle sont recommandés ; la candidate ou le candidat est alors amené-e à commenter ses programmes. Notons que le jury attend en termes d'implémentation plus que de simples appels à des bibliothèques préexistantes (solveur de programmation linéaire, simulateur d'automates cellulaires...).

Le TIPE est un travail d'*intérêt personnel*, ce qui sous-entend une certaine originalité. De nombreux TIPE portent sur des sujets très classiques, proches du programme, ou sur lesquels les sources détaillées et accessibles abondent : algorithmes de plus court chemin sur les graphes, problèmes de flot maximal ou de coupe minimale, l'algorithme min-max ou alpha-beta, etc. De tels sujets doivent être traités de façon *parfaite*, et ne donneront qu'une note moyenne. À l'inverse le jury apprécie fortement le choix de sujets originaux, plus éloignés des sentiers battus.

Pour les enseignantes et enseignants, il est important d'aider les candidates et candidats à circonscrire leur sujet (ni trop étroit, ni trop ambitieux ou trop difficile), de s'assurer qu'elles et ils pourront y développer quelques aspects techniquement riches, et qu'elles et ils ont repéré les références adéquates dans la littérature. Une étape importante est la formalisation des problèmes abordés, et le choix des algorithmes pour les traiter. Un regard critique est important à ce moment là. Le jury a constaté à plusieurs reprises des formalisations maladroites, erronées ou trop lourdes, voire des formalisations de problèmes datées, abandonnées et dépassées par l'état de l'art.

Sans que cela soit des suggestions pour les éditions futures du concours, voici des sujets qui ont été particulièrement appréciés lors de l'édition 2024, pour leur originalité, leur technicité, ou par le traitement particulière pédagogique qui en a été fait :

- Imitation de style musical avec intelligence artificielle
- Le logarithme discret sur une courbe elliptique
- Optimisation de la reconnaissance automatique de cartes à jouer
- Le lemme local de Lovasz, version algorithmique
- Réécriture de graphe comme évaluation β -optimale du λ -calcul
- Formaliser et vérifier les mathématiques avec la théorie des types
- Le principe du raisonnement contrefactuel et ses applications

- Optimisation d'une heuristique avec l'étude dynamique des composantes connexes d'un graphe
- Générations et tests de suites pseudo-aléatoires pour la création d'un jeu de poker
- Le jeu de la vie - Turing-complétude en pratique - L'algorithme Hashlife
- Jeu de Nim en cascade
- Dérivation et comparaison de fonctions définies formellement en informatique
- Décidabilité du problème de ray-tracing
- Simulation de fluides

Écueils à éviter. Certains TIPE se contentent de reprendre un article de recherche très focalisé. Le jury s'attend à ce que l'article soit parfaitement compris, jusque dans ses preuves, à ce qu'il soit remis dans un contexte plus large et positionné par rapport à la littérature sur le sujet, témoignant d'un certain recul. Un travail d'implémentation est aussi attendu.

Le jury a noté depuis quelques années une tendance des candidates et candidats à reprendre des sujets disponibles "clés en main" sur internet, plutôt que de développer leurs propres centres d'intérêt. Ceci n'est pas dans l'esprit de l'épreuve, et comporte de nombreux risques. Outre que cette stratégie est facilement détectée, les sujets sont traités de façon stéréotypée, souvent présentés par plusieurs personnes, et parfois à peine maîtrisés.

Attention aussi aux projets trop ambitieux qui demandent de traiter correctement plusieurs problèmes pour aboutir : cela donne souvent des solutions incomplètes, peu robustes, sans profondeur informatique et qui ne mettent pas en avant un savoir faire pointu. Mieux vaut creuser un des aspects du problème, même si celui-ci n'est qu'une des facettes d'un objectif plus large.

Certains TIPE consistent en des projets de modélisation et simulation. Le jury constate que les réalisations de ce genre font parfois des choix discutables et manquent d'introspection (bien comprendre son modèle et ses propriétés, ses points faibles, le nombre et le rôle des paramètres, etc), par ailleurs elles restent souvent très simples du point de vue algorithmique. La seule simulation d'un modèle (qu'il soit simple ou compliqué) ne peut suffire, si elle n'est pas accompagnée d'une analyse critique : le modèle est-il à la bonne granularité ? Peut-on le simplifier pour produire ou observer les mêmes phénomènes, peut-être plus rapidement ? Surtout, est-ce que les résultats observés peuvent être rapprochés d'une analyse théorique simple (existence d'un comportement limite ou stationnaire, caractérisable, temps de convergence de l'algorithme en moyenne ou au pire cas, etc.) ? Dans la même ligne d'idée, certains sujets visent à simuler des phénomènes physiques, par exemple par intégration numérique d'une équation différentielle. Ils permettent rarement de développer des questions d'informatique.

En 2024, plusieurs candidates et candidats ont opté pour des sujets mettant en jeu des réseaux de neurones. Dans ce cas, au delà de la simple utilisation de bibliothèques (TensorFlow, Keras, ScikitLearn,...) : il faut comprendre et présenter les notions et algorithmes (éventuellement les réimplémenter), savoir argumenter les choix techniques faits (fonctions d'activation, structure des réseaux, nature des données, etc.), démontrer une certaine culture sur le sujet et s'être interrogé-e sur les résultats théoriques disponibles. Le jury s'attend à une analyse critique des résultats et à une démarche scientifique rigoureuse. Ces sujets, dans leur forme la plus simple de réseaux *feed-forward* entraînés par gradient stochastique, laissent peu d'espace à des développements techniques, et peuvent facilement tomber dans les écueils des sujets de type simulation évoqués plus haut.

3 Le rapport

La candidate ou le candidat *doit* rédiger un rapport de synthèse d'au plus 6 pages sur son travail. Des annexes de longueur raisonnable sont autorisées pour présenter tout ou partie du code, ou des preuves détaillées ; elle sont lues à la discrétion des membres du jury.

Le TIPE peut être un travail de groupe, mais, le cas échéant, cela doit être indiqué clairement dans le rapport, avec le nom des autres personnes du groupe. Le rapport reste individuel et doit mettre en évidence la contribution propre de la candidate ou du candidat. Les rapports de TIPE sont systématiquement passés au crible par des outils de correspondance anti-plagiat, et confrontés entre autres aux rapports de l'année courante et des années précédentes.

Le rapport doit replacer le sujet dans son contexte, comporter quelques développements techniques, décrire l'implémentation réalisée, les résultats obtenus. Les schémas et résultats les plus significatifs doivent figurer avant l'annexe. Une bibliographie est attendue, comportant toutes les références utilisées par la candidate ou le candidat.

Le rapport doit avoir une visée pédagogique, démontrant l'esprit de synthèse de la candidate ou du candidat et sa maîtrise de son sujet. Il convient de choisir le bon niveau d'abstraction, de chercher à faire comprendre même des sujets très techniques. Deux écueils sont à éviter : une présentation excessivement technique, noyée sous les notations (voire la paraphrase des références utilisées), ou au contraire une présentation excessivement superficielle, textuelle, rechignant à la formalisation. Il vaut mieux poser les bonnes équations que recourir à un long discours ; on attend une formalisation correcte des problèmes abordés.

4 Déroulement de l'épreuve

L'épreuve consiste en un oral de 40 minutes devant un jury de deux personnes, qui ont lu auparavant le rapport de la candidate ou du candidat. La candidate ou le candidat est tout d'abord invité-e à présenter son travail de manière synthétique. Pour cela, elle ou il peut s'aider d'un support visuel numérique (diaporama format pdf), un ordinateur étant mis à sa disposition. Cette option permet de présenter rapidement des schémas, résultats, courbes, algorithmes, etc., mais d'excellentes candidates et candidats ont pu choisir de faire un exposé « au tableau ». La présentation initiale doit cibler une durée approximative de 10 minutes. Le jury viendra systématiquement interrompre l'exposé pour demander des précisions, tester la culture de la candidate ou du candidat sur son sujet, questionner un choix de modélisation, etc. Avec ce dialogue, l'expérience montre qu'un exposé prévu pour 10 minutes en vient souvent à couvrir toute la durée de l'épreuve. Il est fortement recommandé de répéter en public sa présentation, pour ajuster sa durée, identifier les points techniques les plus intéressants à développer, et s'assurer - le cas échéant - que l'on dispose des bons transparents, ni vides, ni trop chargés, offrant le bon point d'appui aux arguments développés. Cette recommandation est également valable pour les présentations au tableau, qui doivent, pour être réussies, avoir été bien anticipées et préparées.

Bien qu'optionnel, le jury a constaté que disposer d'un support visuel permettait aux candidates et candidats de mieux valoriser leur travail et les aidait à structurer leurs réponses aux questions en servant de point de départ pour la discussion avec le jury. En effet, pour cette épreuve, la candidate ou le candidat est amené-e à répondre aux multiples questions des membres du jury, qui portent soit directement sur le travail présenté (pour éclaircir des points techniques, par exemple), soit sur des sujets connexes (par exemple, des approches alternatives ou des notions de base en informatique). De manière générale, le recul de la candidate ou du candidat sur son sujet et sa culture générale en informatique sont des éléments importants que le jury cherche systématiquement à évaluer à travers des questions sortant du cadre strict de l'étude présentée. Il est à noter que la présentation ne se limite absolument pas à un commentaire du code, mais que le jury cherchera systématiquement à interroger la candidate ou le candidat sur les aspects théoriques de son travail. Voici une liste de situations typiques :

- la candidate ou le candidat utilise une notion complexe et le jury lui propose de la définir précisément ;
- la candidate ou le candidat décrit un algorithme et le jury lui demande d'en évaluer la complexité en temps ou bien de prouver son bon fonctionnement (la preuve du bon fonctionnement de tout algorithme présenté doit pouvoir être donnée par la candidate ou le candidat, ou au moins esquissée dans les grandes

- lignes si celle-ci est difficile ou fait appel à des notions hors programme) ;
- la candidate ou le candidat utilise une structure de données classique et le jury l’interroge sur les algorithmes classiques sur cette structure de données ;
- une candidate ou un candidat expose sa solution au problème qu’elle ou il s’est posé, et le jury le guide vers des structures de données plus performantes pour résoudre son problème.

La candidate ou le candidat peut, à tout moment, consulter ses notes pour s’aider dans ses réponses aux questions du jury. Dans ses interactions le jury s’adapte au niveau de la candidate ou du candidat, ce qui implique que si cette dernière ou ce dernier souhaite utiliser des concepts qui sont hors du programme (par exemple la NP-complétude ou autre), le jury vérifiera systématiquement ses connaissances sur ces points particuliers pour lesquels il est attendu plus qu’une connaissance superficielle.

5 Critères d’appréciation

Le jury appuie son évaluation sur une variété de critères : la profondeur du travail, son ouverture et originalité, les connaissances théorique de la candidate ou du candidat, son aisance et sa pédagogie, la qualité du code, etc. Au delà de la force des résultats obtenus, le jury apprécie la démarche scientifique : le jury valorise la curiosité et la prise de risque ainsi que l’initiative et la ténacité, même si les choix techniques ne sont pas idéaux, ou les résultats pas à la hauteur des objectifs de départ. Certaines candidates et certains candidats ont eu la note maximale, non pas parce que leurs travaux étaient parfaits ou exempt de toute critique, mais pour récompenser un TIPE remarquable sur plusieurs critères (originalité, prise de risque, curiosité, maîtrise des éléments techniques, rigueur de la démarche, bonne exploration de la littérature...).

La qualité de la documentation scientifique est cruciale, à commencer par une recherche bibliographique autour du sujet choisi : on attend une bonne maîtrise des sources et quelques résultats centraux (incluant leurs preuves, si elles sont simples, ou à défaut leur principe), ainsi qu’une exploration des méthodes alternatives. Les interprétations empiriques ou reprenant des éléments de presse scientifique « grand public » sont à proscrire.

Le jury attend un sens critique dans l’analyse des résultats. Les candidates et candidat sont encouragés à s’interroger sur pourquoi une approche fonctionne ou ne fonctionne pas, à expliquer les phénomènes observés, à les vérifier avec une contre-expérience, etc. Pourrait-on faire aussi bien avec une méthode plus simple ? Quel est l’effet de tel ou tel paramètre ? Et ainsi de suite. Les démarches expérimentales sont souvent intéressantes : elles peuvent donner lieu à l’acquisition de données de terrain, la réalisation de dispositifs, etc. Le jury note cependant qu’elles mènent souvent les candidats à « bricoler » une méthode, sans en extraire des aspects fondamentaux pour l’informatique. Il est important d’isoler un problème exprimé rigoureusement, de rechercher les solutions classiques dans la littérature, de les comparer, d’évaluer les résultats avec des outils pertinents... Plutôt que de partir de zéro, il est préférable de s’appuyer sur des articles spécialisés quitte à suivre un protocole expérimental existant.

Pour ce qui est des réalisations logicielles, l’« emprunt de code » doit être systématiquement indiqué. Les candidats doivent prendre soin de distinguer le code emprunté de celui produit, et de dire dans quelle mesure celui-ci est inspiré de code disponible en ligne. Il convient d’expliquer les choix techniques (langage de programmation, bibliothèques logicielles utilisées). Attention à savoir distinguer la description générale d’un algorithme et son implémentation : les choix d’implémentation méritent une argumentation surtout s’ils exploitent astucieusement des propriétés de l’algorithme et de la machine pour gagner en performances.