

Banque MP Inter-ENS - Session 2022

Rapport sur l'épreuve de TIPE d'informatique

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

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

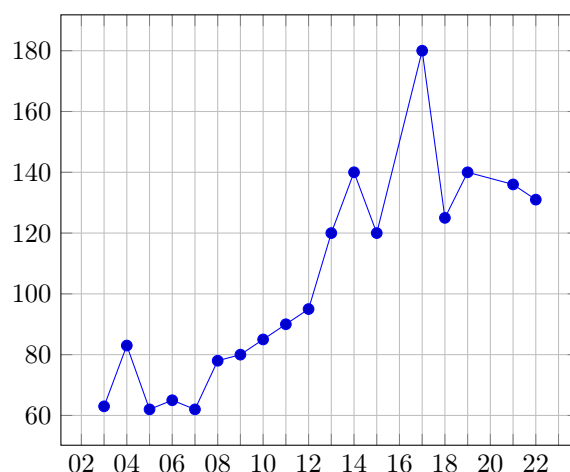
	ENS Paris-Saclay	ENS Lyon	ENS (Paris)	ENS Rennes
Concours MP :	3.8%	5.4%	7.4%	3.8%
Concours Info :	7.9%	5.6%	3.3%	11.4%

Membres du jury : P. CLAIRAMBAULT, É. FABRE, M. RAO, M. SIGHIREANU

1 Éléments de statistiques

Le jury a évalué cette année 131 candidats présentant un dossier TIPE d'informatique. Le nombre de candidates (4 cette année) reste anormalement faible au regard de la proportion des chercheuses en informatique (supérieure à 25%). Le jury encourage vivement les enseignants à diriger les candidates vers des sujets d'informatique. Il s'efforce de rester mixte, et demeure très vigilant sur les biais de genre dans son appréciation des sujets traités.

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 aussi bien dans le thème de l'année qu'en dehors. Les candidats ont été évalués sur leur maîtrise des concepts informatiques généraux, sur la qualité scientifique de leur travail, et surtout sur leur maîtrise et leur compréhension du sujet

choisi. Le jury tient à souligner que le niveau des candidats auditionnés s'est significativement amélioré depuis une dizaine d'années. De très nombreux oraux étaient de niveau très satisfaisant et plusieurs de niveau excellent ; de nombreux candidats se sont révélés lors des questions. Le jury a particulièrement apprécié l'enthousiasme, la créativité et la curiosité de certains candidats, et 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, et le jury n'a pas hésité à mettre plusieurs fois la note maximale.

2 Choix du sujet

Le choix du sujet est particulièrement important et doit permettre au candidat de mettre en valeur ses capacités créatives, la rigueur de son approche, son esprit critique et sa capacité à s'impliquer dans un projet qui le motive. L'originalité et la prise de risque sont souvent valorisées par le jury, surtout quand elles traduisent une implication particulière des 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, une domaine applicatif, etc. Chaque fois que le sujet s'y prête, un travail expérimental et une réalisation logicielle sont attendus ; le candidat est alors amené 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-tend une certaine originalité. De nombreux candidats choisissent 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, 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 encadrants : il est important d'aider les candidats à circonscrire leur sujet (ni trop étroit, ni trop ambitieux ou trop difficile), de s'assurer qu'ils pourront y développer quelques aspects techniques riches, à leur portée, et qu'ils ont repéré les bonnes références 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, bricolées ou trop lourdes, voire des formalisations de problèmes anciennes, abandonnées et dépassées par l'état de l'art. De même, il a pu noter l'utilisation d'algorithmes approchés alors que des solutions plus performantes ou exactes étaient disponibles.

Sans que cela soit une recommandation pour les éditions futures du concours (le jury changeant chaque année), les sujets suivants ont été particulièrement appréciés lors de l'édition 2022, par leur originalité, leur technicité, ou par le traitement particulièrement pédagogique qui en a été fait :

- la clusterisation de graphes
- la théorie algorithmique des noeuds, son application au dénouage de l'ADN
- la modélisation du repliement des protéines et son traitement algorithmique
- l'indécidabilité du problème du mot (appartenance d'un mot à un groupe défini par sa présentation)
- l'exploration de questions de calculabilité pour le jeu d'échecs sur plateau infini
- le fonctionnement des solveurs SAT modernes
- le calcul efficace d'invariants en théorie des noeuds
- ...

Écueils à éviter. Certains candidats 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 du candidat. Un travail d’implémentation est aussi attendu.

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

Beaucoup de candidats choisissent 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 memes 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 liées proprement à l’informatique.

En 2022, beaucoup de candidats ont opté pour des sujets mettant en jeu des réseaux de neurones. Le jury s’attend à plus que, par exemple, 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 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

Le candidat doit rédiger un rapport de synthèse d’au plus 6 pages sur son travail.

Le TIPE peut être un travail de groupe, mais, le cas échéant, cela doit être indiqué dans le rapport, ainsi que le nom des autres personnes ayant travaillé dans le groupe. Le rapport reste individuel et doit mettre en évidence la contribution propre du candidat. Les rapports de TIPE sont systématiquement passés au crible par des outils de correspondance anti-plagiat, et confrontés entre autres aux rapports 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 être dans le corps du texte. Une bibliographie est attendue, comportant toutes les références utilisées par le candidat. Une annexe est possible, notamment pour les résultats secondaires (courbes, schémas, images,...), des preuves indispensables mais trop longues, et pour les éléments significatifs du code. On a pu observer un usage abusif des annexes, et des rapports allant de 12 à 20 pages, avec le risque qu’ils ne soient pas entièrement lus en détail.

Le style télégraphique ou la forme d’aide-mémoire technique sont à bannir, de même que la paraphrase d’articles scientifiques. Le rapport doit au contraire avoir une visée pédagogique, démontrant l’esprit de synthèse du candidat et la 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.

Lors de l'édition 2022, le jury a pu déplorer des rapports remis en retard, manquants, incomplets, ou remplacés par un diaporama d'exposé. Il en a été tenu compte dans la notation.

4 Déroulement de l'épreuve

L'épreuve consiste en un oral de 40 minutes devant un jury de 2 personnes, qui ont lu auparavant le rapport du candidat. Le candidat est tout d'abord invité à présenter son travail de manière synthétique. Pour cela, 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'excellents 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 du candidat sur son sujet, questionner un choix de modélisation, etc. 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.

Bien qu'optionnel, le jury a constaté que disposer d'un support visuel permettait aux 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, le candidat est amené 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 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. Voici une liste de situations typiques : le candidat utilise une notion complexe et le jury lui propose de la définir précisément ; le candidat décrit un algorithme et le jury lui demande d'en évaluer la complexité en temps ou bien de prouver son bon fonctionnement ¹ ; le candidat utilise une structure de données classique et le jury l'interroge sur les algorithmes classiques sur cette structure de données ; un candidat expose sa solution au problème qu'il s'est posé, et le jury le guide vers des structures de données plus performantes pour résoudre son problème. 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 du candidat, ce qui implique que si ce dernier souhaite utiliser des concepts qui sont hors du programme (par exemple la NP-complétude ou autre), le jury vérifiera systématiquement les connaissances du candidat 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 du candidat, son aisance et pédagogie, la qualité du code, etc. Au delà de

1. La preuve du bon fonctionnement de tout algorithme présenté doit pouvoir être donnée par le candidat (ou au moins esquissée dans les grandes lignes si celle-ci est difficile ou fait appel à des notions hors programme).

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. Certains candidats ont eu la note maximale, non pas parce que les travaux étaient parfaits ou non discriminables, 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 candidats 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 bien distinguer l'algorithmique des choix d'implémentation : ceux-ci ne sont pertinents que s'ils permettent d'exploiter astucieusement des propriétés de l'algorithme, ou d'en réduire la complexité.