
Inattention, intelligence artificielle autonome et responsabilité

Yves Oytana^{*†1} and Marie Obidzinski²

¹Centre de REcherches sur les Stratégies Economiques (UR 3190) (CRESE) – Université de Franche-Comté, Université de Franche-Comté : UR3190 – IUT 30 Avenue de l’observatoire 25009 Besançon Cedex, France

²Centre de Recherche en Economie et Droit (CRED) – Université Paris-Panthéon-Assas – Centre Valette, 19-21 rue Valette / Rez-de-jardin, 75005 Paris, France

Résumé

Nous caractérisons la règle de partage de responsabilité socialement optimale dans une situation où un concepteur développe une intelligence artificielle (IA) à destination d’utilisateurs humains. Le concepteur choisit un niveau d’investissement dans le développement de l’autonomie de l’IA et fixe un prix de vente. L’utilisateur choisit ensuite d’acquiescer ou non l’IA. Même si l’IA est utilisée, son autonomie reste limitée, nécessitant l’intervention de l’utilisateur dans certaines situations. Les utilisateurs sont sujets à de l’inattention comportementale, réduisant l’efficacité de leurs interventions (une mauvaise intervention crée un dommage). En conséquence, il est supposé que l’IA est plus performante que l’utilisateur humain. De plus, certains utilisateurs ne sont pas conscients d’être sujets à de l’inattention comportementale. Nous mettons en évidence l’existence d’un arbitrage lors du choix du partage de responsabilité. Sous certaines conditions, le concepteur investit insuffisamment dans le développement de l’autonomie de l’IA. Le décideur public peut inciter ce dernier à augmenter son investissement en augmentant la part du dommage qu’il supporte. D’un autre côté, une telle hausse de la responsabilité du concepteur peut freiner la diffusion de l’IA, réduisant ainsi le surplus social.

Mots-Clés: règles de responsabilité, intelligence artificielle, inattention, prédiction, algorithme performatif

*Intervenant

†Auteur correspondant: yves.oytana@univ-fcomte.fr