Heterogeneous Interactions in Artificial Intelligence Development Races

T. Cimpeanu[1], *T.A. Han*[1], *F. C. Santos*[2,5], *L.M. Pereira*[3], *T. Lenaerts*[4,5] *(1) Teesside University, Middlesbrough, UK, t.cimpeanu@tees.ac.uk (2) Universidade de Lisboa, Porto Salvo, Portugal (3) Universidade NOVA de Lisboa, Caparica, Portugal (4) Vrije Universiteit Brussel, Brussels, Belgium (5) Université Libre de Bruxelles, Brussels, Belgium*

The field of Artificial Intelligence (AI) is introducing a certain level of anxiety in research, business, and policy making. Tensions are further heightened by an AI race narrative which makes many stakeholders fear that they might be missing out. Whether real or not, a belief in this narrative may be detrimental, as some stakeholders will feel obliged to cut corners on safety precautions or ignore societal consequences just to "win". Starting from a baseline model describing an idealised technology race in a well-mixed world, we investigated how different spatial structures underlying the network of contacts among the race participants can alter the evolutionary outcomes and requirements for regulatory actions. Our findings indicate that, when participants portray a strong diversity in terms of connections and peer-influence (e.g., when scale-free networks shape interactions among parties), the conflicts that exist in homogeneous settings are significantly reduced, thereby lessening the requirement for regulatory actions.

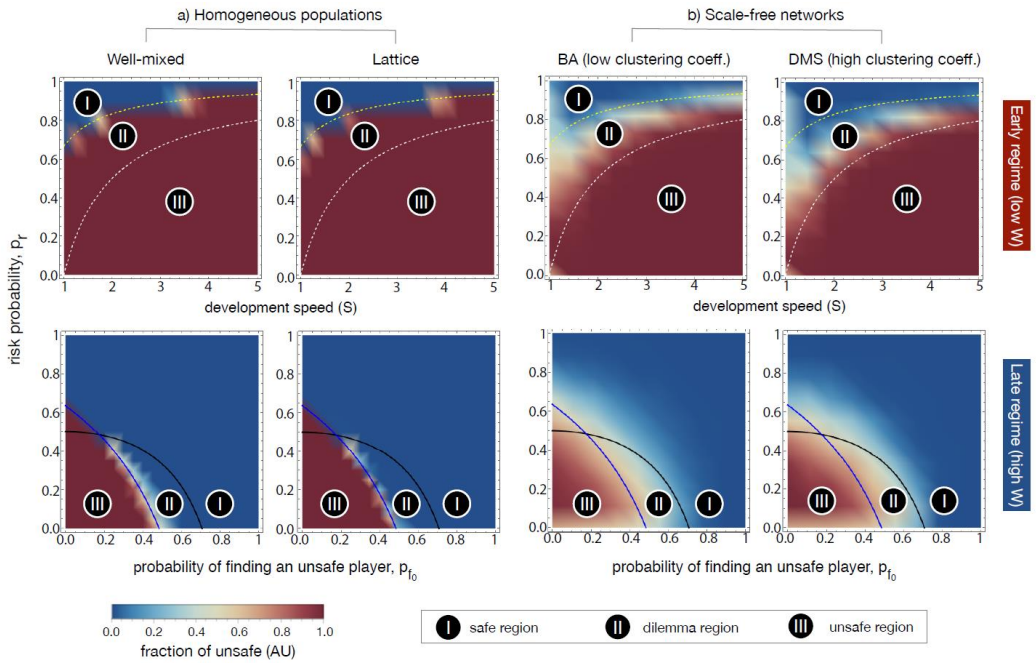

Figure 1: Color gradients indicating the average fraction of AU (unsafe strategy) for different networks and regimes. Dotted and full lines indicate the phase diagram regions obtained analytically. Parameters: $p_{f_0} = 0.5$ and $W = 100$ (top panels); $s = 1.5$ and $W = 10^6$ (bottom panels); $c = 1$, $b = 4$, $B = 10^4$, $\beta = 1$, in all panels.

References

[1] T. Cimpeanu, F. C. Santos, L. M. Pereira, T. Lenaerts, T. A. Han. (2020). "AI Development Race Can Be Mediated on Heterogeneous Networks (working title)". *Preprint Journal Submission.* URL: https://arxiv.org/abs/2012.15234.