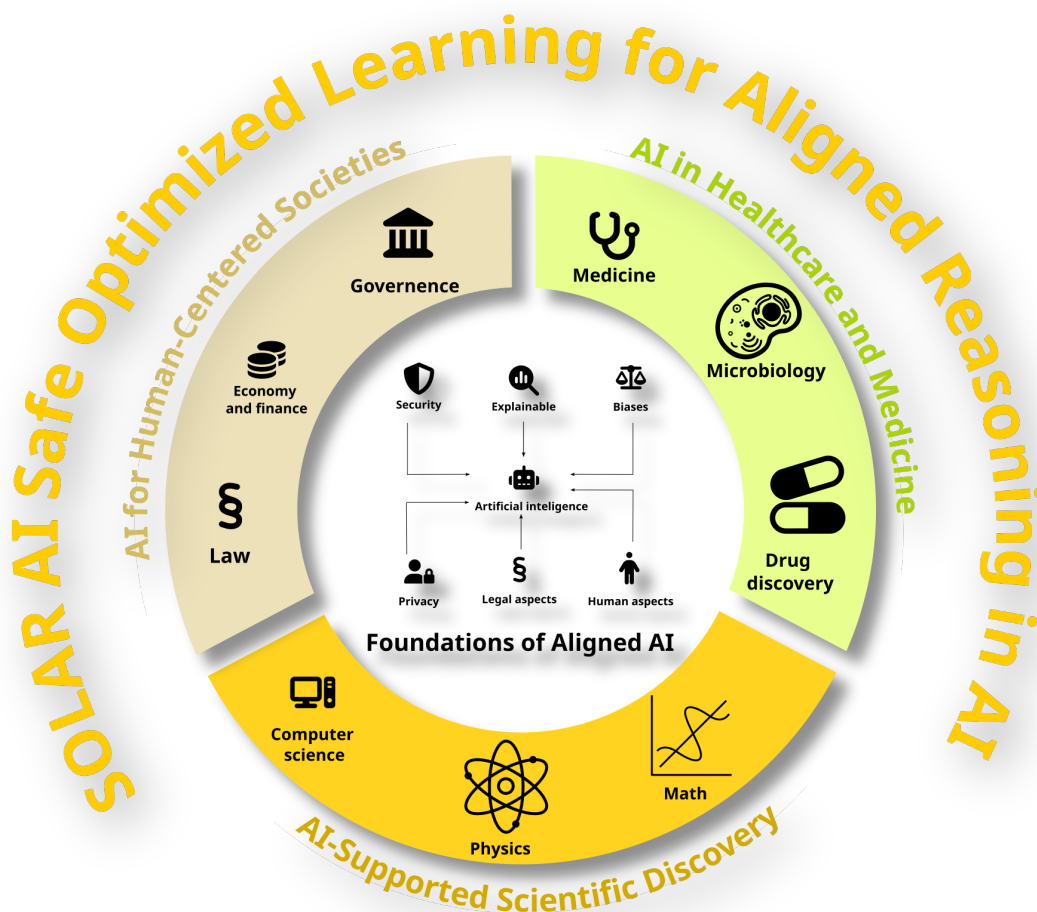


SOLAR-AI (Safe Optimized Learning for Aligned Reasoning in AI) to interdyscyplinarny projekt badawczy, którego celem jest rozwój bezpiecznych, interpretowalnych systemów sztucznej inteligencji zgodnie z ustalonymi wartościami w społeczeństwie, nauce i medycynie. Projekt zakłada, że dostosowanie AI do wartości i norm zależy od kontekstu, a oczekiwania wobec systemów różnią się w tych trzech obszarach.



SOLAR-AI łączy badania podstawowe z wdrożeniami, tworząc rozwiązania wspierające podejmowanie decyzji i przyspieszające odkrycia naukowe, jednocześnie dbając o zaufanie, bezpieczeństwo i etykę. Rozwijane są metody uczenia wartości, wyjaśnialności, ograniczania uprzedzeń i podejmowania decyzji w warunkach niepewności, które stanowią fundament praktycznych zastosowań.

W społeczeństwie AI wspiera decyzje w obszarach takich jak prawo, finanse czy zarządzanie, zapewniając interpretowalność i zgodność z normami prawnymi oraz etycznymi. W nauce, np. w fizyce, matematyce czy informatyce, wspomaga modelowanie i tworzenie hipotez, dbając o poprawność i replikowalność wyników. W medycynie działa bezpiecznie, przejrzysto i rozliczalnie, zgodnie z zasadami etyki i standardami medycznymi, zwłaszcza w decyzjach dotyczących zdrowia i godności pacjentów.

SOLAR-AI dostarcza ramy tworzenia systemów AI, które uwzględniają specyficzne wartości i ryzyka właściwe dla poszczególnych dziedzin.