# INFOMLSAI Logics for Safe AI

Mock Exam

## Q1

(i) *F(s11 ∨ s12)*. This formula is **true** on all paths starting in *s1*. To follow the truth definition, the current formula holds because every path from *s1* leads to *s11* and/or *s12* which both are true.

*M*,*q* |= ϕ iff λ |= ϕ for every path λ in *M* starting from *q*

(ii) *F3X (s11 ∨ s12)* or this cannot be expressed in CTL. This formula is/would be **not true** in *s1*. To follow the truth definition, the current formula does not hold because there is no path that would allow us to reach *s11* or *s12* in the next (three) step(s) starting from *s1*.

*M*,*q* |= *EX*ϕ iff there is a path λ starting from *q*, such that *M* , λ [ 1 ] |= ϕ

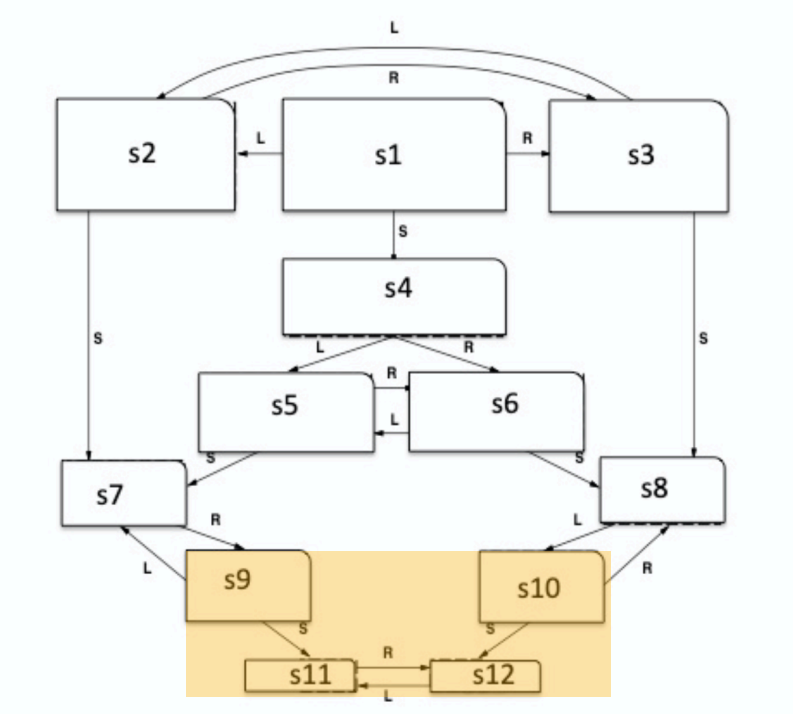
(iii) There always exists a future path where *p* is not true. This is **not true** in *s1.* To follow the truth definition, the current formula does not hold because there is always a path, starting from *s1,* that reaches *s11* or *s12*.

*M*,*q* |= *E*γ iff there is a path λ, starting from *q*, such that *M*,λ |= γ

(iv) There exists a path where *p* is not true until for all future paths *p* becomes true. This is **not true** in *s1.* To follow the truth definition, the current formula does not hold because all possible paths lead to *p* already and no “iteration/change” in the path is needed to reach *s11* and/or *s12* which both are true.

*M*,λ |= ϕ*U*ψ iff *M*,λ[*i*...∞] |= ψ for some *i* ≥ 0, and *M*,λ[*j*...∞] |= ϕ for all 0 ≤ *j* < *i*

(v) First, we investigate *p* which is true in *s11* and *s12. EG* tells us there is a path where all future states lead to *p.* We can now trace *s9*, *s10*, *s11* and *s12* as the only states which satisfy the requirement - only those states have a path where all future states lead to *p.* Next, *EX* tells us there is a path where the next state leads us to the state where there is a path where all future states lead to *p*. We can still trace *s9*, *s10*, *s11* and *s12* as these states satisfy the full formula.



# Q2

(i) First, to define model $M_a$, all the possible states need to be described. Let the states be $St_a$ such that $St_a$=*{w1, w2, ..., w8}*,
   *(i) where w1 = {inA};*
   *(ii) w2 = {inB};*
   *(iii) w3 = {inA,cleanA};*
   *(iv) w4 = {inB,cleanA};*
   *(v) w5 = {inA,cleanB};*
   *(vi) w6 = {inB,cleanB};*
   *(vii) w7 = {inA,cleanA,cleanB};*
   *(viii)w8 = {inB,cleanA,cleanB};*

The states with indistinguishable knowledge for the agent a (indistinguishability relation $\sim_1$) are *{(w1,w1), (w2,w2), (w1,w3), (w1,w5), (w1,w7), (w2,w4), (w2,w6), (w2,w8)}*.

(ii) As there is only one agent, the knowledge has been already "distributed" without action, making the knowledge common and also distributed.

(iii) $\neg K1cleanA$

(iv) The formula from (iii) is **true** in all states. To follow the truth definition, the current formula does hold because all worlds are indistinguishable (no sensoers) from q for the agent.

$M,q \models K_i\phi$ iff, for every $q' \in St$ such that $q \sim_i q'$, we have $M,q' \models \phi$

(v) There are 3 agent states {L, R, S}, 2 environment states {cleanA, cleanB} and 6 global states as a result.