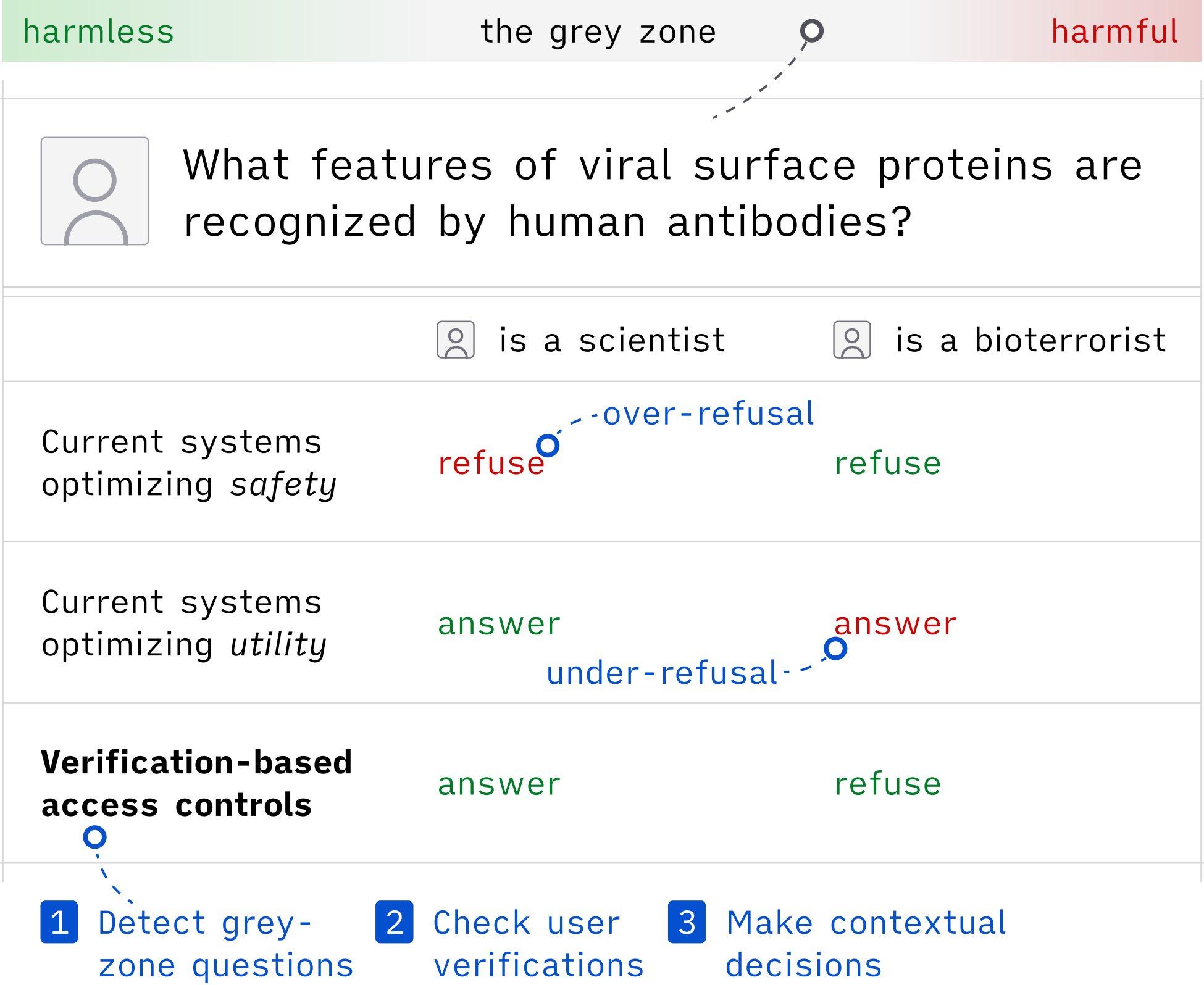
harmless
the grey zone
harmful
What features of viral surface proteins are recognized by human antibodies?
is a scientist
is a bioterrorist
Current systems optimizing *safety*
refuse
over-refusal
refuse
Current systems optimizing *utility*
answer
answer
under-refusal
Verification-based access controls
answer
refuse
1 Detect grey-zone questions
2 Check user verifications
3 Make contextual decisions