

危险与风险的区别

1 安全是一个概率过程

世界上不存在绝对的安全。安全的定义是“不存在不可接受的风险”，而不是不发生事故。

和疾病问题一样，安全是一个概率过程，某些健康问题和交通事故都是永远无法避免的。安全设计的目标是把安全变成可控，控制在“可接受范围之内”。如果安全系统的设计理念是“避免一切事故”，那么结果一定是失败，因为，安全是一个概率过程，是概率驱动，而不是一个确定过程，不可用微分方程来描述。把概率过程当成确定过程来管理，必然无法阻止小概率事件（长尾/corner 事件）的发生。这是当前安全理论的一个主要局限性。所以需要寻找新的评估方法来论证与评价，这便是安全大脑的方向：降低危害发生的概率。

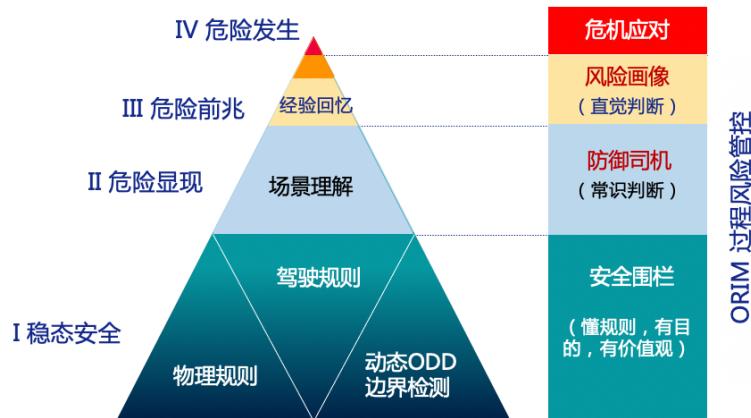
目前 AV 安全技术主要是寻求必然性解决途径，把安全当成确定过程来处理，而忽略了概率性。稳健的安全策略应当将安全划分为应对危机（确定性过程）和风险管理（概率性过程）两阶段分阶段管理，也就是先期用右脑进行风险过滤，临近事故时采用左脑进行确定性过程危机应对（而不是像理想那样采用反过程）。

在传统的自动驾驶安全技术里，大多不划分这两类性质的挑战，采用统一的应对措施，直到危险临近时才采取紧急应对措施，对系统性能提出了极高的要求，也增加了应对失败的概率。

安全设计的大目标应是追求大概率，采用什么技术路线都可以。IPM 次数、透明性、可解释性本身并不是目的，而只是通向“大概率安全”的手段。

我们需要相信概率的力量。其中有两层含义：一是相信在数据统计层面上，概率一定会呈现相应的趋势规律，在大数据面前，事物一定会按照相应的概率分布，这意味着首先要相信风险金字塔是一个事实。二是我们努力要做的是提高有利事件的发生概率，降低有害事件的概率。这就是安全金字塔的内涵。

风险金字塔理论见[1].



2 危险和风险应当分别控制

风险和危机不是一个发生机理，一个是确定性过程，另外一个是概率过程，所以，适用于这两个目的的世界模型是不同的。

临近事故的金字塔顶端适于用“物理世界模型”来描述（图 3），因为事件的发生是确定性的，动作目标是确定的，物理世界模型的任务是评价这个事件和动作是否安全。

金字塔的下部风险是不确定发生的危险（图 3），是否发生取决于个体之间的互动关系，也就是社会关系。虽然风险不一定发生，但是风险要素是确定的，风险-危险的比例也是确定的，因此人们制定了一系列的风险方案措施，比如交规和防御驾驶规则。降低风险的发生率等同降低危险的发生率。风险的描述更适合于用“社会世界模型”。

物理模型是一个完全的理性世界，没有弹性和灰区，这种思维方式与人类左脑接近。社会模型注重定性的关系分析，具备常识，理解约定俗成，运用直觉判断，思维方式与人类右脑接近。所以完整的世界模型应当由左脑-右脑两个部分构成的。这种特征我们称之为安全世界模型的“双模特征”。

右脑是人脑不可或缺的组成部分。人有健全的双脑时，左脑负责局部，右脑负责整体态势。左脑侧重于控制有规律的日常行为，右脑专门负责处理威胁等这类非日常情况的反应。比如，当生物遇到捕食者要吃掉自己时，需要立即采取适当的行动保证生命安全，而右脑就是为了处理这些紧急事务进化而来的。从这种意义而言，右脑就是安全大脑。

如果失去了右脑 Right Hemisphere Brain Damage (RHD)，人类会

- 失去整体概念，没有社会关联感，无法与人交往
- 没有注意力分配机制，所以注意力涣散，时刻分心，无法专注，没有关注中心，容易受扰动，所以执行效率非常低下，要么过激，要么迟缓，会表现得很神经质
- 无法推理和理解语言暗示，无法理解笑话，无法理解局势
- 没有长期记忆，难以回忆重要的经验，也无法学习，无法安排计划

- 没有目的性，不知道问题的存在，更不知道寻找解决问题的方法
- 无法感知左侧信息

开发安全右脑的目的在于：

- 识别风险和避险
- 提高左脑执行任务的效率

另外，右脑的思维方式与 M Minsky 提出的六层意识模型是非常接近的（图 2）[2]，要求具备在不同层级进行多层抽象的能力。可为将来进入到自主任务分解打下基础。自主任务分解要求机器要分层思维，能够把陌生的任务分解成一系列熟悉的子任务组合，再用子任务的答案综合出陌生任务的解决方案。

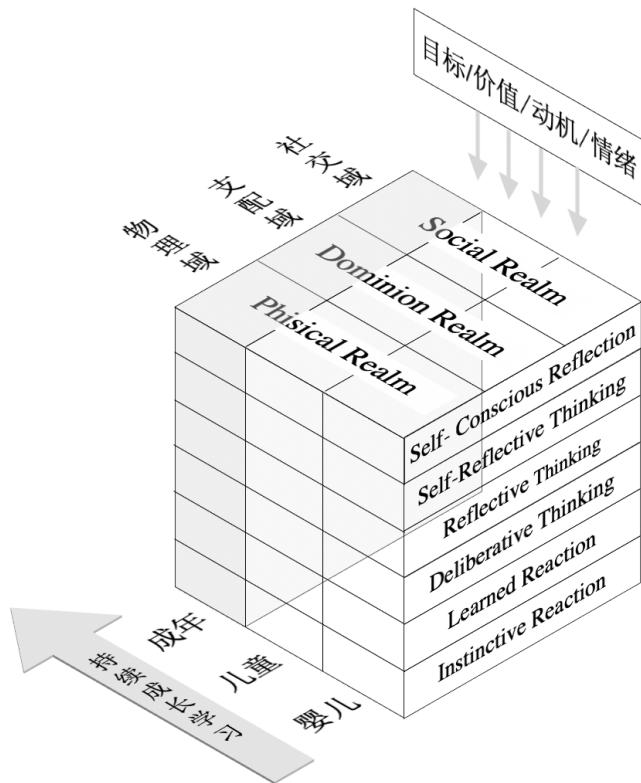


图 2 六层 AV 意识模型：思维异构化是人类智慧（resourcefulness）的来源

上述的内容要求（物理世界+社会现实）、思维方式要求（左脑+右脑）、安全的分类控制概念（风险+危险）之间的对应关系可统一表达成图 3。

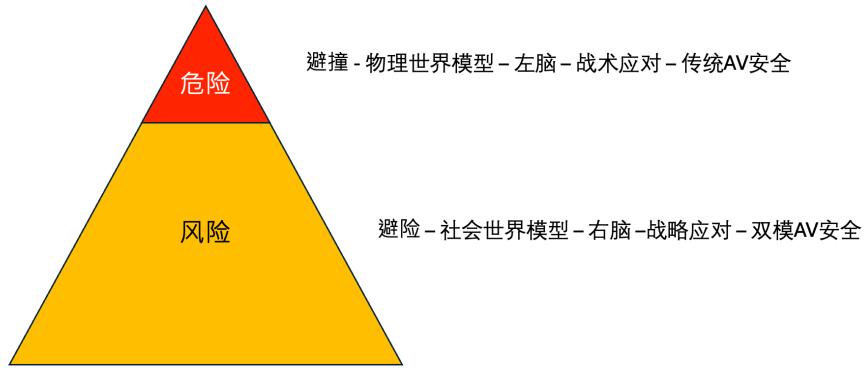


图 3 风险 vs 危险、左脑 vs 右脑、物理模型 vs 社会模型、战术 vs 战略

参考文献

- [1] 风险控制金字塔理论
- [2] Minsky 意识模型与 AV 世界模型