

大数据热中的冷思考

当前，由于大数据在社会实践与理论研究上的巨大影响，有关大数据的讨论已经渗透到每一个行业和领域之中，带来了各行各业对大数据的热捧。大数据显然具有重要价值，这一点毋庸置疑。正如有学者指出的，大数据开启了一次重大的时代转型，

它正在改变我们的生活以及理解世界的方式，成为新发明和新服务的源泉。因此，我们要大力实施国家大数据战略。但在大数据热中也应该有冷思考，因为大数据可能带来的风险与挑战是不容忽视的。只有正视各种风险与挑战，才能更好推动大数据发展。

防止大数据的“去人性化”忽视人的主体价值和数据的社会文化意义。大数据的一个根本缺陷就是它只关注“客观数据”的呈现，而缺少对“主观数据”的反映。从社会学角度看，所有数据说到底都是有关“人”的符号，因此，符号所蕴含的社会文化意义是我们真正需要了解和掌握的内容。如果脱离了人及其所在社会的价值本性，无限放大大数据的效应及其影响力，就可能本末倒置，最终产生许多难以预期的后果。数据来源于社会又作用于社会。无论是可以数字化的数据还是不可数字化的数据，其比较、分析和归纳其实反映的都是人在符号层面上的一种互动。大数据是汇集了行为主体、时间、地点三个维度的“数据宇宙”。在大数据时代，个人的信息更多的是以数据为载体来进行互动的。如果人及其各类社会行为都仅仅被看作一个个数据符号，我们就很难从数据中解读出充满朝气、富有激情的人，也就很难从数据塑造出来的没有个性特征的人去推导群体样态和社会构成。

防止大数据的“巨量化”掩盖数据个性特征、强化数据霸权。犹如网络一样，大数据也是一把双刃剑。巨量化的大数据在给我们研究和分析社会现象带来便利的同时，实际上也掩盖了许多具有个性特征的数据的意义。在连续的巨量数据海洋中，许多本身具有个性特征的数据已经变得无足轻重了，它们在大数据霸权面前会成为微不足道的牺牲品，大数据简化了人们对数据差异性的认知。大数据的复杂性不仅在于其数据样本巨量，更在于其多源异构、多实体和多空间之间的交互动态性。当前，只有少数人掌握处理这种复杂的巨量化大数据的技术，这很容易导致“数据暴力”。从这个意义上说，防止大数据滋生“数据暴力”是大数据发展中需要注意的问题。

防止大数据的“碎片化”导致人们难以系统认知复杂事物。从某种意义上说，大数据的价值只有在其能够被规范成为可供分析的形式之后才能最大程度地被挖掘出来。然而，在巨量的大数据面前，实际上可供规范分析的数据只是极少数，许多数据都停留在“碎片化”阶段而难以被真正挖掘和分析。对这些碎片化的数据进行清理，形成规整的结构，是大数据能够得到充分利用的根本所在。在大数据中，其研究范式重在发现而不是推论，运用大数据进行各种分析不应像传统的社会调查方法那样通过假设检验来进行推论，而应通过数据的总体归纳来达成对社会现象的总体分析。这不仅要求我们在数据清理技术、存储结构上进行优化，而且要在数据采集环节尽可能地做到规范，否则就很难获得对复杂事物的系统认知。

防止大数据的“模糊化”产生大量信息垃圾、为数据造假与不当传播打开方便之门。在大数据时代，我们真正需要的并不仅仅是大数据本身，而是大数据背后所隐藏的我们想要获取的各种信息资源，而这些信息资源可能只占数据总体的万分之一甚至更少。从这个角度说，大量的信息其实都是垃圾。不仅如此，许多大数据本身就是模糊的，其中含有大量虚假和有害的内容。如果我们纯粹凭借数据来判断和分析人类行为，甚至不假思索地去利用和传播某种数据，就会导致许多误判。因此，未经前期调研论证和规范分析所获取的信息，其数据量越大可能越模糊，其得到的垃圾信息也就可能越多，这也为各种数据造假和不当传播打开了方便之门。对此，我们必须学会辨识和判断数据的真实性，避免因盲目的数据崇拜造成误判，甚至迷失在大数据的洪流之中。