# Analizando la Integración Humano-Sistema en Sistemas Autónomos

Miriam Gil, Manoli Albert, Joan Fons y Vicente Pelechano

Centro de Investigación PROS, Universitat Politècnica de València
Camí de Vera s/n, 46022 Valencia, Spain
{mgil, malbert, jjfons, pele}@pros.upv.es

**Resumen.** Los sistemas autónomos (SA) están diseñados para actuar de forma autónoma en gran parte de su trabajo; sin embargo, la autonomía completa es una utopía a medio y corto plazo. Este hecho hace necesario que el humano ayude a completar su funcionalidad ("human-in-the-loop"). Este tipo de sistemas deben garantizar en todo momento un correcto funcionamiento autónomo, a la vez que debe ceder, bajo ciertas condiciones, total o parcialmente el control al humano para la realización de algunas tareas. Esto requiere analizar y diseñar los sistemas para que involucren al humano de forma adecuada ante situaciones donde no es posible alcanzar la autonomía, procurando garantizar una correcta integración humano-sistema. En este trabajo se proporcionan las bases para analizar y diseñar las interacciones humano-sistema. En este artículo se presenta un análisis que permite identificar los aspectos esenciales de la participación del humano en el SA y se propone una técnica para especificar cómo integrar el humano y el sistema en las primeras fases de desarrollo. Los coches autónomos se toman como ejemplo para ilustrar la propuesta mediante escenarios reales.

**Palabras clave:** Sistemas Autónomos, "Human-in-the-Loop", Interacción Humano-Sistema, Coches Autónomos

## 1 Introducción

Los sistemas autónomos (SA) deben poder actuar de forma autónoma para llevar a cabo las tareas para las que fueron concebidos. Sin embargo, la variedad de sistemas, dominios, entornos y contextos de ejecución junto con las restricciones tecnológicas, legales y sociales actuales impiden una completa autonomía a corto/medio plazo. Este hecho hace necesaria la introducción del humano para complementar la funcionalidad del sistema (lo que se conoce como *"human-in-the-loop"*). Según [11], el escenario de la autonomía completa de los SA no llegará en mucho tiempo. Mientras tanto los SA requerirán soporte humano para poder garantizar un funcionamiento correcto en todas las situaciones. El papel que juega el humano en los SA dependerá del nivel de autonomía del SA, y determinará el rol que debe jugar el humano en el control y ejecución de una tarea. Este rol puede ir desde la responsabilidad sobre la monitorización, planificación, o ejecución de la tarea, hasta la supervisión de la misma [13].

En el contexto de la computación autónoma uno de los bucles de control más aplicados para monitorizar y regular los sistemas autónomos es el bucle MAPE-K [16]. Este bucle se compone de cuatro actividades principales: monitorizar, analizar, planear y ejecutar. Las soluciones *human-in-the-loop* introducen al humano en el bucle de control

de forma que éste pueda participar en las distintas fases: monitorización (ayudando al sistema a identificar situaciones del contexto), planificación (colaborando con el sistema en la toma de decisiones) y ejecución (manejando, dirigiendo o ejecutando tareas). Estas soluciones son complejas de desarrollar porque integrar los sistemas con los humanos no es una tarea trivial; los humanos deben ser capaces de cooperar con los sistemas de forma eficiente e intuitiva. La integración humano-sistema [5] debe garantizar en todo momento un correcto funcionamiento autónomo (desde un punto de vista operacional, de calidad de servicio, de seguridad, etc.), y a su vez debe ceder, bajo ciertas condiciones, el control total o parcial al humano para llevar a cabo algunas tareas. El SA debe evaluar hasta qué punto el humano es capaz de ejecutar las tareas que el sistema requiere, dentro de unas restricciones espacio-temporales (entre otras), y configurar un marco de interacción adecuado para ceder el control de dichas tareas al humano.

Para diseñar la interacción humano-sistema es necesaria una nueva visión de la relación entre el humano y el sistema, ya que se persigue una completa integración de ambas partes [5]. Por ejemplo, en el caso de los coches autónomos, el humano necesita el coche para viajar cómodamente sin estar pendiente de la conducción, y el coche necesita al humano para gestionar situaciones que no puede resolver por sí solo. Esta integración debe ser natural, no intrusiva y robusta, en la que, además de los aspectos funcionales y autónomos del sistema, se tengan en cuenta las intenciones, el estado, las emociones y las acciones del humano que se puedan inferir. La presencia y el comportamiento del humano pasa a ser un factor clave del sistema. Siguiendo con el ejemplo del coche autónomo, es necesario maximizar la probabilidad de que el humano tome el control del coche cuando éste lo requiera (colaboración robusta); pero también aspiramos a un coche que sepa captar la atención del humano dependiendo del grado de implicación que requiera la tarea, evitando una demanda excesiva de atención (colaboración no intrusiva). Considerar el diseño de la participación del humano un tema menor o relegar su tratamiento únicamente a la fase de implementación puede llevar a SA que no se integran bien con los humanos.

Este trabajo pretende sentar las bases que permitan analizar y diseñar la integración del humano en los SA de forma que se satisfagan las necesidades del humano y del sistema. Nuestro objetivo es proporcionar herramientas conceptuales que ayuden al diseñador a concebir y analizar, en etapas tempranas del desarrollo, las tareas que humano y sistema deben realizar de forma colaborativa. Estas herramientas permitirán a los diseñadores entender y describir *dónde* y *porqué* surge la necesidad de involucrar al humano, *cuándo* se puede llevar a cabo la tarea de colaboración (condiciones que deben cumplirse), *qué* acciones son necesarias, *cómo* se lleva a cabo la interacción y *quién* debe realizarla (perfil de humano requerido). Proponemos, además, un modelo de ejecución que define cuál será el modo de funcionamiento de la integración humano-sistema y cómo el SA cederá parcialmente el control al humano para llevar a cabo las tareas de colaboración. El diseñador podrá utilizar estas herramientas para especificar (mediante plantillas que proponemos) las características esenciales y el comportamiento (a través del modelo de ejecución) de las tareas de colaboración entre el humano y el sistema. Esta documentación inicial servirá de entrada a la *fase de diseño*, donde se transformará la especificación de las tareas de colaboración en interfaces, y mecanismos de interacción concretos que se diseñarán en función de las particularidades de cada sistema autónomo, y se identificarán y modificarán aquellos componentes software o servicios del sistema autónomo

que deben extenderse para tener en cuenta el nuevo comportamiento del sistema a la hora de involucrar al humano.

El resto del trabajo se estructura de la siguiente forma. La Sección 2 analiza los trabajos relacionados que abordan la interacción humano-sistema en sistemas autónomos. La Sección 3 realiza una caracterización de la participación del humano en el SA, donde se identifican los principios que establecen cómo debe ser esta participación, y se analizan los aspectos esenciales de la misma. En la Sección 4 se propone una técnica basada en el uso de plantillas para describir las características esenciales de las tareas de colaboración entre humano y SA. Se define también un modelo de ejecución que operacionaliza la participación del humano en el SA. En la Sección 5 se presenta un caso de estudio en el dominio de los coches autónomos para ilustrar la propuesta. Finalmente, la Sección 6 presenta el trabajo futuro y las conclusiones.

## 2 Trabajos Relacionados

Durante mucho tiempo se han realizado estudios para definir qué tipo de interacciones deberían existir entre humanos y sistemas que automatizan tareas [6]. Esta automatización desplaza el rol que juega el humano sobre el control de una tarea desde un rol de responsabilidad de planificación, ejecución y monitorización, a un rol relacionado con la supervisión de la tarea [13]. Tal y como se plantea en [20], los SA pueden ofrecer mejoras sustanciales, como incrementar la seguridad, reducir el error humano y disminuir los tiempos de respuesta, entre otros. Sin embargo, antes que estos sistemas sean una realidad, hay que resolver otros retos relacionados con factores humanos como el efecto negativo causado por adaptaciones provocadas por malos usos, una dependencia excesiva sobre los procesos automáticos o incluso cambios en el nivel de atención (distracciones) de los humanos. En [11] se argumenta que vamos a entrar gradualmente en el mundo de los SA en el que tendremos sistemas con diferentes niveles de autonomía. La autonomía completa se define como la capacidad de los sistemas de actuar, bajo cualquier circunstancia y para todas sus capacidades, de manera autónoma. Según [11], este escenario no llegará en muchos años. Mientras tanto los SA requerirán soporte humano para poder garantizar un funcionamiento correcto en todas las situaciones.

Otros trabajos relacionados han intentado identificar problemas vinculados con las interacciones humanas en los SA. Por ejemplo, en [9, 18] abordan el problema de la desconfianza que genera al humano delegar responsabilidades de control en los sistemas. Promueven el desarrollo de SA que mantengan informado, den retroalimentación continua sobre los procesos subyacentes, descarguen carga mental y estrés, a la vez que mantengan en el humano cierta "consciencia situacional" (representación mental) sobre estos procesos. Argumentan la necesidad de diseñar soluciones que satisfagan estos requisitos para aumentar las posibilidades de que los humanos acepten estos sistemas.

Varios autores han destacado la necesidad de la participación del usuario en la resolución de conflictos de sistemas autónomos [19, 2]. En [12] se presenta un estudio del papel del humano en los sistemas ciber-físicos. Otras propuestas como [1, 4] proporcionan una solución parcial (centrándose en tareas de ejecución) para razonar acerca de la integración del usuario en los procesos de auto-adaptación. Sin embargo, ninguna de ellas proporciona mecanismos ni técnicas para especificar cómo debería ser esa integra-

ción y participación del humano. Tal como se afirma en [5], las características de los nuevos sistemas software obligan a un cambio en la forma de entender la interacción humano-sistema. En los SA, humano y sistema colaboran de forma conjunta en la realización de ciertas tareas. Es necesario cambiar el foco en el diseño de las interacciones hacia una concepción holística de la colaboración entre el humano-sistema, para llegar a sistemas que integren al humano de forma natural.

Del análisis de estos trabajos se desprende que la participación del humano en los SA es un campo que debe explorarse y abordarse con más profundidad para desarrollar correctamente estos sistemas. En [14] los autores identificaron los retos tecnológicos que introduce la participación del humano en los SA y definieron un marco conceptual para caracterizar la cooperación humano-SA. En este trabajo se extiende la propuesta mediante técnicas para el análisis y diseño de la participación del humano en las primeras etapas de concepción de los SA.

## 3 Principios y Aspectos Esenciales de la Participación del Humano en los SA

Asumimos que los SA que integran a los humanos se componen, entre otros, por un conjunto de tareas, procesos y actividades automáticas que permiten desempeñar la función para la que han sido concebidos. El sistema es responsable en todo momento de identificar qué tareas debe realizar. Para ello, debe monitorizar el entorno en el que opera para activar estos procesos automáticos [16]. El sistema deber ser capaz de identificar situaciones en las que no se pueda solucionar de manera completa, correcta o con garantías una determinada tarea. En este caso el sistema debería hacer partícipe al humano para que, de manera colaborativa, resuelvan la tarea. Este trabajo colaborativo implica que, aunque el humano tenga el control de cierta tarea, el sistema debe supervisar las acciones del humano para comprobar que éste está ejecutando las acciones que se esperan y garantizar el éxito de la colaboración.

La participación del humano en los SA introduce un considerable número de retos a resolver para conseguir una integración humano-sistema correcta y sin fisuras, bien integrada en el SA incluso en situaciones donde los usuarios tengan limitados los recursos atencionales, cognitivos o físicos para llevar a cabo la interacción. Estos retos los capturamos a través de tres principios básicos (identificados en la literatura como retos de interacción entre humanos y sistemas autónomos [22, 23]) que reflejan cómo debe ser la participación del humano en los SA. Los tres principios son los siguientes:

- **Principio 1. Interacción Natural.** El humano debe involucrarse para llevar a cabo o ejecutar las tareas que el SA no puede llevar a cabo por sí solo. El SA debe ofrecer al humano los mecanismos de interacción y la información adecuada para poder ejecutar la funcionalidad requerida. Lo que se pretende es maximizar la posibilidad de que el SA ejecute con éxito su funcionalidad.
- **Principio 2. Gestión del Nivel de Atención.** El SA debe ser capaz de captar la atención del humano (el humano puede no tener puesta la atención en el sistema) pero a su vez debe evitar molestarle en exceso. Esto implica que es necesario realizar una gestión específica de los recursos de atención del humano. El SA deberá captar su atención dependiendo del grado de implicación que sea necesario, evitan-

do abrumarlo con interacciones innecesarias o acciones que demanden excesiva atención y destreza (reducir la molestia). Es importante establecer un equilibrio entre el nivel de autonomía y la intervención humana para evitar llegar a resultados no deseados o una mala experiencia de usuario.

- **Principio 3. Transparencia y retroalimentación.** El SA debe disponer de los mecanismos necesarios para ofrecer retroalimentación de forma que el humano comprenda el funcionamiento del SA y confíe en él. Los humanos puede que no confíen en los SA que actúan sin "explicar" o proporcionar información sobre el contexto de ejecución autónoma o las decisiones que toman. Si no saben qué va a suceder en el sistema o cómo va a reaccionar este ante una determinada situación, el comportamiento autónomo puede provocar desconfianza [9].

Estos principios establecen cómo debe ser la participación del humano en el SA. Tomando estos principios como base, en la siguiente sección analizamos la participación del humano en el SA con el objetivo de identificar los factores que influyen en el diseño de esta participación para conseguir una correcta integración humano-sistema.

### 3.1 Análisis de la participación del humano en el SA

En el ámbito de los sistemas auto-adaptativos, las *six honest serving men questions* (when, where, what, who, how, why) han tenido especial relevancia en la literatura como método para el análisis de problemas [8 y 16]. En [16] se utilizan para analizar y elicitar los requisitos esenciales de la auto-adaptación. En este trabajo proponemos la utilización de estas cuestiones para identificar los aspectos que nos permitan caracterizar la participación del humano en los SA.

**Where:** ¿dónde participa el humano?. Es necesario localizar la "*situación*" que requiere involucrar al humano. En el caso de los SA el *dónde* será la **tarea** (del sistema) que requiere la participación del humano. Por ejemplo, en los coches autónomos, el coche, circulando por autopista en modo autónomo ante una entrada a ciudad, pide al humano que tome el control (según los niveles de SAE[1] [17] en el nivel 3 el coche no puede conducir de forma autónoma por ciudad). En este caso el humano ayudaría a ejecutar la tarea de conducción. Las tareas que el humano puede realizar en el SA pueden ser de diferentes tipos. Teniendo en cuenta diferentes aproximaciones [3, 9, 20, 21], hemos clasificado las tareas que se requiere del humano, y proponemos los siguientes grupos:

- *Tareas de ejecución*, el humano será responsable de manejar, ejecutar o dirigir la tarea (asistido o no por el sistema). Por ejemplo: *tomar el control del coche.*
- *Tareas de decisión*, el humano debe decidir entre un conjunto de opciones predefinidas por el sistema, o completamente por su cuenta. Por ejemplo: *confirmar cambio de ruta por accidente en la ruta habitual*; o *decidir si pasar a modo manual y reducir el consumo ante batería baja.*
- *Tareas de monitorización*, el humano proporcionará información al sistema acerca del contexto difícil de monitorizar. Por ejemplo: *Informar al coche de una incidencia en la carretera o de la señalización de un policía*.

**When**: ¿cuándo participa el humano?, sabemos dónde debe participar (*where*) pero no siempre podrá hacerlo. Nos debemos preguntar ¿qué condiciones se deben cumplir para

[1] La SAE (Sociedad de Ingenieros Automotrices) clasifica los coches en 6 niveles de autonomía (de 0 a 5) desde los sistemas más básicos a la conducción 100% autónoma.

que el humano pueda y/o sea capaz de participar en el sistema? Se deberá tener en cuenta el entorno del sistema, la situación del humano y el estado del propio sistema. Las condiciones a considerar sobre estos aspectos dependen de la perceptibilidad del sistema, que determina qué estados puede captar el SA. Estas condiciones de contexto determinan si es factible la participación del humano en la tarea en la situación de contexto actual (*precondición*). Por ejemplo, para realizar la tarea *Tomar el control del coche* es necesario que "haya un humano en el coche con capacidad de conducción", "sentado en el asiento de conducción" y "que esté atento a la conducción". Si esto se cumple, el humano está preparado para realizar la tarea. Si no se cumple la precondición se debe reflexionar sobre si ¿es posible realizar acciones que lleven a cambiar el contexto y satisfacer la precondición? O cuestionarnos si ¿existen situaciones que hagan imposible satisfacer la precondición y, por tanto, imposibiliten que el humano participe en el sistema? Por ejemplo, la condición "un humano sentado en el asiento de conducción" es condición necesaria para participar en la tarea, pero que no se cumpla no imposibilita la ejecución de la misma. Esto es porque se podría requerir que el humano realizara un cambio de asiento, y si este lo hiciera ya estaría preparado para participar en la tarea. Lo mismo ocurre con la condición "que esté atento a la conducción"; se puede alertar al usuario para que pase a estar atento. Sin embargo, la "no presencia en el coche de un humano con capacidad de conducción", imposibilitaría la realización de la tarea, ya que no es posible cambiar esa situación para llegar a satisfacer la precondición. Por lo que esta sí sería una condición que imposibilitaría la realización de la tarea. Es decir, podemos identificar condiciones sobre la precondición cuyo no cumplimiento imposibilitaría la ejecución de la tarea (*condiciones inhabilitantes*). Respecto al resto de condiciones de la precondición que no son inhabilitantes, si no se satisfacen en el momento de lanzar la tarea, el sistema podría realizar acciones para tratar de cambiar el contexto y que logren satisfacerse, cumpliendo así el *principio 2*.

**What**: ¿qué debe hacer el humano para participar en el sistema? ¿qué secuencia de acciones seguirá la interacción humano-sistema? Se debe desgranar la tarea en el conjunto de acciones que la componen. Las acciones serán responsabilidad o bien del humano o bien del sistema. Se distinguen diferentes tipos de acciones:

- *Acciones propias de la tarea*, acciones necesarias para realizar la tarea. Deben permitir una adecuada simbiosis entre el humano y el sistema, ofreciendo al humano los mecanismos de interacción y la información adecuada para poder ejecutar la funcionalidad requerida (*principio 1)*. Por ejemplo, para la tarea *Tomar el control del coche* serían: (1) el sistema informa al humano de que tome el control y (2) el humano coge el volante.
- *Acciones de retroalimentación*, se corresponden con acciones donde el sistema informa al humano. Ofrecen información para que el humano comprenda el funcionamiento del SA y deposite su confianza en él (*principio 3)*. Por ejemplo, en la tarea *Tomar el control del coche* una vez el humano ha cogido el volante: (3) el sistema confirma al humano que le ha pasado el control del coche.
- *Acciones de preparación*, correspondientes a acciones para alcanzar las condiciones adecuadas que permitan al humano hacer su tarea, es decir, satisfacer la precondición (*when*). Están orientadas a captar la atención del humano (*principio 2).* Siguiendo con la tarea *Tomar el control del coche,* si el humano con capacidad de conducción no estuviera sentado en el asiento de conductor, un paso de preparación

sería: (0) avisar al humano de que se sitúe en el asiento de conducción para disponerse a tomar el control del coche.

**Who**: ¿quién es el humano que va a participar en el sistema?, ¿cuál es su perfil? No todos los perfiles de humanos podrán participar de la misma forma. Para cada tarea, será necesario caracterizar qué perfil de humano (indicando sus capacidades) es necesario para poder llevar a cabo de manera satisfactoria dicha tarea. En el caso de la tarea *Tomar el control del coche*, el perfil del humano debe ser un pasajero mayor de 18 años.

**How**: ¿cómo participa el humano en el sistema? A nivel de análisis, relacionamos esta cuestión con el grado de atención que necesitamos del humano (dependiendo del nivel de atención se elegirán unos mecanismos de interacción u otros) y con las restricciones que pueden existir sobre su participación, ya que ambos factores influyen para conseguir una integración óptima humano-sistema. Respecto al nivel de atención, teniéndolo en cuenta conseguiremos un diseño que consiga captar la atención del humano cuando sea necesario, pero a su vez evite molestarle en exceso (*principio 2)*. El nivel de atención lo asociaremos a las acciones que componen una tarea. Hemos establecido tres niveles de atención básicos: alto, medio y bajo (que podrían refinarse en función del escenario concreto). Cuando se realice el diseño de un sistema concreto, los niveles de atención determinarán el tipo de mecanismo de interacción específico que se debe utilizar para realizar la interacción con el humano [7]. Por ejemplo, en el caso del coche autónomo el paso de preparación "*avisar al humano de que se sitúe en el asiento de conducción para disponerse a tomar el control del coche*" requiere un nivel de atención alto. Esto se podría traducir en que el sistema realizaría la petición a través de los altavoces y de un mensaje en la consola de comunicación del coche. Mientras que el paso de retroalimentación "*el sistema confirma al humano que éste tiene el control del coche*" tiene un nivel de atención bajo por no ser un mensaje tan crítico. Esto se traduciría en que el sistema realizaría la confirmación únicamente con un mensaje en la consola del coche. Respecto a las restricciones, las tareas se realizarán cumpliendo una serie de condiciones (temporales, contextuales, etc.). Nos debemos preguntar ¿qué condiciones se deben cumplir durante la realización de la tarea para asegurar que esta se realiza de forma exitosa? Estas condiciones serán típicamente temporales (por ejemplo, el humano debe seleccionar una opción en menos de 10 segundos), aunque también pueden ser de otro tipo (como restricciones de espacio, por ejemplo, el humano tiene que seleccionar una opción antes de que el coche recorra 500 metros). El sistema comprobará en todo momento que las restricciones se cumplen. Si las restricciones no se cumplen se ejecutarán acciones de emergencia que dejen al sistema en un estado consistente y seguro.

**Why**: ¿por qué tiene que participar el humano? En un sistema autónomo la participación (en mayor o menor grado) del humano vendrá determinada por la imposibilidad del sistema de realizar por sí solo una tarea. Esto puede ocurrir ante situaciones en las que el sistema se encuentra en un contexto desconocido o no tiene una acción programada para resolverlo. Siguiendo con el coche autónomo, según la clasificación de niveles de SAE, si el coche está realizando una conducción autónoma en el nivel 3 por vía rápida sin retención, y se encuentra próximo a entrar a una ciudad, requerirá que el humano realice la tarea *Tomar el control del coche*.

En esta sección se han definido los principios que determinan cómo debe ser la participación del humano, y se han identificado una serie de aspectos esenciales que se deben

especificar para diseñar la participación del humano en el SA. En la siguiente sección se utilizan estos principios y aspectos esenciales como base para definir:

a) Una técnica para describir las características de la participación del humano en el SA. Esta técnica ofrece una notación para capturar formalmente los aspectos esenciales de la participación del humano identificados a través de las preguntas.
b) Un modelo de ejecución que determina cómo se llevará a cabo a nivel operacional la participación del humano en el SA. Este modelo define el modo de funcionamiento de la integración humano-sistema y cómo el SA cederá parcialmente el control al humano para llevar a cabo las tareas de colaboración.

# 4 Caracterizando la Integración Humano-Sistema

En esta sección se propone, en primer lugar, una herramienta conceptual que permitirá capturar la información relevante para analizar la participación del humano en el SA (identificada en la sección anterior). El objetivo es permitir al diseñador de la integración humano-sistema especificar las necesidades, condiciones y restricciones de esta integración, y razonar sobre ellas, descubriendo problemas de interacción humano-sistema antes de que el sistema sea construido (sección 4.1). En segundo lugar, se propone un modelo de ejecución que permitirá operacionalizar la integración humano-sistema. Este modelo de ejecución detalla los pasos, las restricciones, y el orden entre los pasos, para ejecutar las tareas de colaboración. El modelo de ejecución sirve al diseñador para saber cómo se deben llevar a cabo estas tareas (sección 4.2).

## 4.1 Especificación de la Integración Humano-SA

Las *six honest serving men questions* estudiadas nos han servido para determinar un conjunto de aspectos esenciales de la participación del humano en el SA. Estos aspectos permiten identificar la información que se necesita capturar para el análisis de la integración humano-sistema. A partir de este estudio, hemos elaborado un marco de trabajo que, a través de una serie de preguntas, intenta elicitar para cada una de las características identificadas, los factores a considerar en el diseño de la integración humano-SA. La Tabla 1 recoge este marco de trabajo.

| Característica | Preguntas | Factores a considerar |
|---|---|---|
| Evento o Condición | ¿Por qué se requiere al humano? | Evento o condición que requiere que el humano participe en una tarea colaborativa. |
| Actividad | ¿Dónde participa el humano? ¿qué tarea realiza el humano? | Actividad que debe realizar el humano para ayudar al SA a complementar su funcionalidad. El papel del humano puede ser: *Decisor*, *Actuador* o *Informador*. |
| Actor | ¿Qué perfil de humano se requiere para ejecutar la tarea? | Quiénes son los humanos que participan en la interacción y cuáles son las capacidades que se le requieren (perfil). |
| Precondición | ¿Qué condiciones se deben cumplir para que el humano pueda parti- | Condiciones sobre el propio sistema, su entorno y el humano (según la perceptibilidad del sistema) que determinan que el humano esté en la situación reque- |

| | | |
|---|---|---|
| | cipar en la tarea? | rida para realizar la tarea. |
| Condiciones inhabilitantes | ¿Qué condiciones impiden que el humano participe en la tarea? | Condiciones sobre el propio sistema, su entorno y el humano que imposibilitan que se satisfaga la precondición necesaria para realizar la tarea. |
| Restricciones | ¿Existen restricciones en la ejecución de la tarea? | Condiciones que se deben cumplir durante la ejecución de la tarea para garantizar el éxito de la misma al finalizar la participación del humano. |
| Pasos | ¿Qué secuencia de pasos sigue la interacción humano-sistema? ¿cuál es el nivel de atención requerido? | Pasos de interacción humano-sistema que componen la comunicación necesaria para la ejecución de la tarea. Los pasos se pueden asociar con restricciones y con el nivel de atención requerido en cada paso. El nivel de atención se categoriza en: alto, medio o bajo. |

**Tabla 1.** Características de las tareas de integración humano-SA.

Para capturar toda esta información, proponemos una plantilla que describe las tareas de colaboración (ver Tabla 2). El objetivo es que un diseñador de la interacción pueda centrarse en definir la información relevante para conseguir una efectiva integración.

| | | |
|---|---|---|
| Nombre Tarea | Nombre o identificador de la tarea | |
| Evento o Condición | Condiciones de contexto o eventos ocurridos que dispararán la participación del humano | |
| Descripción | Descripción de la tarea donde participa el humano | |
| Actor | Descripción del perfil de humano que participa en la interacción y descripción de las capacidades que se le requieren | |
| Precondición | Condiciones sobre el propio sistema, su entorno y el humano requeridas para que el humano pueda realizar la tarea | |
| Condiciones inhabilitantes | Condiciones sobre el propio sistema, su entorno y el humano que inhabilitan que el humano pueda participar en el sistema, por no poder satisfacer la precondición | |
| Acciones para cambiar el contexto | Pasos que realiza el sistema para intentar alcanzar el cumplimiento de las precondiciones | |
| Pasos (relación temporal de arriba a abajo) | | |
| Humano<br>Acciones o interacciones del humano<br>[nivel de atención] | | Sistema<br>Acciones o interacciones del sistema<br>[nivel de atención] |
| Restricciones | Condiciones temporales, o según dominio podrían ser espaciales o de otros tipos, que se deben garantizar durante la ejecución de la tarea | |
| Plan de Emergencia | Acciones que realizará el sistema como plan de emergencia. | |

**Tabla 2.** Plantilla de elicitación de características de interacción humano-sistema

### 4.2 Modelo de Ejecución para la Integración Humano-SA

Además de permitirnos identificar la información a extraer para especificar las tareas de colaboración, el análisis de las *six honest serving men questions* nos sirve también para determinar cómo se llevará a cabo a nivel operacional la participación del humano en el SA. Es decir, cuál será el modo de funcionamiento y cómo el SA cederá temporal y par-

cialmente el control al humano para llevar a cabo las tareas de colaboración. El modo de funcionamiento describe la secuencia de pasos que hay que llevar a cabo durante la ejecución de una tarea de colaboración (como la comprobación de condiciones, la validación de restricciones, o la realización de acciones de preparación del humano para involucrarlo en una tarea), y el orden o relación entre los pasos. Este conjunto de pasos forma parte de lo que sería el modelo de ejecución de la colaboración humano-sistema.

El funcionamiento del SA (basado en bucles de control autónomos) debe rediseñarse para integrar al humano en el sistema y llevar a cabo tareas de forma colaborativa. El SA debe operar de manera autónoma monitorizando y analizando el contexto, planificando adaptaciones en base al contexto y ejecutando esas adaptaciones. Durante ese proceso puede detectar una situación que dispare la necesidad de involucrar al humano en una tarea de colaboración. En ese momento, el sistema activará la tarea cediendo la responsabilidad al usuario de ejecutar la tarea (de manera colaborativa). La tarea se podrá realizar siempre que se cumplan unas condiciones de contexto (sobre el usuario, el entorno y el sistema); si no se cumplen las condiciones se activará un plan de emergencia que permita dejar al sistema en un estado consistente y seguro. Cuando el problema ha sido resuelto, el sistema vuelve a su comportamiento autónomo (hasta el próximo conflicto en la autonomía).

A partir de este funcionamiento extendido del SA, las cuestiones analizadas nos ayudan a definir el modelo de ejecución. El modelo de ejecución se puede abstraer con los siguientes pasos (ver diagrama de flujo en la Figura 1):

1) Elevar un evento que dispara la necesidad de que el humano lleve a cabo una tarea.
2) Comprobar la existencia de un humano que cumpla el perfil requerido.
3) Comprobar la precondición y las condiciones inhabilitantes (condiciones sobre el contexto que determinan si existen las condiciones necesarias para que el humano participe en el SA):
    a. Si existen las condiciones necesarias: ir al **paso 4**.
    b. Si no existen las condiciones necesarias, pero no se cumplen las condiciones inhabilitantes: ir al **paso 5**.
    c. Si no existen las condiciones necesarias, y se cumplen las condiciones inhabilitantes: ir al **paso 6**.
4) Realizar la tarea: el SA facilitará que el humano lleve a cabo la tarea mediante los mecanismos de interacción y control necesarios. Ir al **paso 7**.
5) El sistema realizará acciones para cambiar el contexto con el objetivo de alcanzar las condiciones necesarias de participación del humano. Volver al **paso 3**.
6) Ejecutar plan de emergencia que deje al sistema en un estado seguro. Ir al **paso 7.**
7) Una vez finalizada la tarea el sistema pasará automáticamente a modo autónomo (hasta detectar nuevas situaciones que requieran la participación del humano).

El sistema comprobará en todo momento (pasos 3, 4 y 5) que las restricciones de la tarea se cumplen hasta su finalización (representado con líneas azules discontinuas en la Figura 1). Si las restricciones no se cumplen se ejecutará el paso 6.

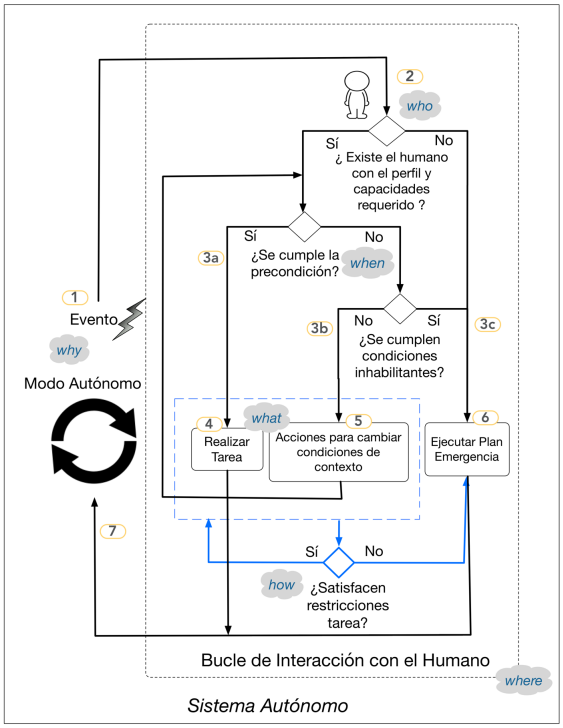


**Figura 1.** Modelo de ejecución.

## 5 Caso de Estudio. El Coche Autónomo

Uno de los sistemas autónomos más de moda en la actualidad son los coches autónomos. Las funciones de la conducción autónoma se incorporarán por fases en los próximos años, pero el conductor seguirá al volante, ya que el máximo nivel de autonomía (sin volante ni mandos) es un nivel todavía utópico [15]. El ejemplo que hemos elegido para ilustrar nuestra propuesta se enmarca dentro de esta problemática. Suponemos un coche autónomo con un nivel de autonomía 3, como por ejemplo el coche de Google [10]. El coche dispone de un servicio que permite la conducción autónoma por vías rápidas (controlando límites de velocidad, distancia de seguridad, maniobras de adelantamiento, etc.). Este servicio opera bajo ciertas condiciones de contexto (como el tipo de vía rápida o condiciones meteorológicas idóneas), por lo que requiere la interacción del humano en casos donde no se cumplan estas condiciones. El nivel 3 nos ofrece también un mecanismo de "fallback" que en caso de necesitar al humano para tomar el control y éste no estar disponible ofrece asistencia al coche para conseguir una situación de riesgo mínimo. Por tanto, este coche posee algunas limitaciones que requieren supervisión y control humanos para lograr una funcionalidad completa. A continuación, se describen varios escenarios de participación del humano.

**Escenario 1: Tomar control del coche**

El coche realiza una conducción autónoma de nivel 3, cuando se aproxima a una ciudad (este coche no puede conducir en modo autónomo por ciudad). Para pasar a modo

manual, se establece un espacio de seguridad de 500 metros antes del cual el humano debe tomar el control. El sistema debe asegurar que el humano pone su atención en el coche para tomar el control, por lo que el nivel de molestia de las interacciones debe ser alto. Pueden darse 2 situaciones: el humano toma el control en el periodo de seguridad establecido (<500 metros) o no lo hace. En el caso de que no lo haga, se activará el plan de emergencia: el coche realiza una maniobra en la que busca un escenario de riesgo mínimo para desactivar la función autónoma y ceder el control al humano (como aparcar en la cuneta con luces de emergencia). La siguiente plantilla especifica este escenario:

| Nombre Tarea | **Tomar Control del Coche** | |
|---|---|---|
| Evento o Condición | El coche se encuentra en un nivel de autonomía 3 circulando en modo autónomo por vía rápida y se aproxima a entrar en ciudad. | |
| Descripción | El humano debe tomar el control del coche -> Tarea de control | |
| Actor | Humano en el interior del coche mayor de 18 años | |
| Precondición | Humano con capacidad de conducción en el coche, sentado en el asiento del conductor y atento | |
| Condiciones inhabilitantes | Ningún humano con capacidad de conducción | |
| Acciones para cambiar el contexto | [Si no hay humano ocupando el asiento del conductor] Avisa al humano con capacidad de conducción de que se tiene que sentar para conducir *[molestia:alta]*<br>[Si el humano no está atento] Alerta para captar su atención *[molestia:alta]* | |
| **Pasos** | | |
| **Humano** | | **Sistema** |
| **2**. Coge el volante | | 1. Pide al humano que coja el volante *[molestia:alta]*<br>3. Pasa a modo manual<br>4. Informa al humano sobre la situación del coche |
| Restricciones | "Distancia recorrida" < 500 metros | |
| Plan de Emergencia | Activar el mecanismo de "fallback" para llevar al coche a una situación de riesgo mínimo, según el contexto (por ejemplo, estacionar el coche en la cuneta con las luces de emergencia) | |

**Escenario 2: Resolver conflicto en gestión energética - Batería Baja**

El coche realiza una conducción autónoma y detecta un nivel de batería bajo (< 20%), pero todavía no crítico (> 5%). Podría seguir ofreciendo todos los servicios de conducción autónoma (durante un tiempo), o tratar de reducir el consumo energético desactivando algunos servicios, ampliando la autonomía de la batería. Se plantea un diseño en el que se delegue en el humano (con capacidad de conducción) la decisión de si el vehículo debe desactivar algunos servicios para ahorrar energía. Esta interacción se realizaría siempre que dicho conductor no esté ocupado en otra tarea, puesto que en estos momentos no es una decisión crítica (podemos inferir pues un nivel medio de molestia de las interacciones). En caso de que la interacción no pudiera darse, o no se diera dentro de un contexto seguro (por ejemplo, el nivel de batería debe ser siempre > al 5%), esta interacción se abortaría. No se requiere un plan de emergencia para este escenario, ya que mientras el conductor no indique lo contrario, el vehículo seguirá circulando en modo autónomo. La siguiente plantilla muestra la especificación de este escenario.

| Nombre | **Resolver conflicto en gestión energética - Batería Baja** | |
|---|---|---|
| Evento o Condición | El coche se encuentra en modo autónomo y se detecta que la batería pasa a nivel bajo ("Nivel batería" < 20%), pero no crítica ("Nivel batería" > 5%) | |
| Descripción | El humano debe decidir si seguir con la conducción autónoma o pasar a modo manual para consumir menos energía -> Tarea decisión | |
| Actor | Perfil pasajero mayor de 18 años | |
| Precondición | Humano con capacidad de conducción | |
| Condiciones inhabilitantes | Ningún humano con capacidad de conducción | |
| Acciones para cambiar el contexto | No son necesarias | |
| **Pasos** | | |
| **Humano** | | **Sistema** |
| 3. Selecciona una de las dos opciones[2] | | 1. Avisa al conductor de la situación *[molestia: media]*<br>2. Da a elegir al usuario si quiere seguir en modo autónomo o cambiar a manual *[molestia: media]*<br>4. Informa al humano del resultado |
| Restricciones | Mientras el "Nivel de batería" > 5% | |
| Plan Emergencia | Continuar en modo autónomo | |

# 6 Conclusiones

En este trabajo se ha analizado qué información se debe tener en cuenta para especificar la participación del humano en el ámbito de los SA. En primer lugar, se han introducido una serie de principios que ayudan a garantizar la efectividad de la integración humano-sistema, y que se deben tener en cuenta cuando se aborda el diseño de la participación del humano en el SA. En segundo lugar, a través de las *six honest serving men questions* hemos realizado un análisis que permite determinar qué aspectos se deben considerar para diseñar la participación del humano teniendo en cuenta los principios marcados. La identificación de los aspectos esenciales de la participación del humano en SA nos ha permitido definir una técnica basada en plantillas para la especificación de la interacción humano-sistema en SA. Los diseñadores pueden utilizar esta técnica en las primeras fases del desarrollo del sistema para entender la interacción humano-sistema y capturar los requisitos de interacción para su posterior diseño. Como trabajo futuro se pretende validar la plantilla propuesta comprobando si el uso de dichas plantillas facilita el diseño de estos sistemas. Otra línea de trabajo futuro es la definición de técnicas, lenguajes o notaciones para diseñar la interacción humano-sistema a partir de la especificación capturada mediante las plantillas propuestas.

**Agradecimientos.** Trabajo financiado por el MINECO bajo el proyecto PROTEUS TIN2017-84094-R, y cofinanciado por la *Generalitat Valenciana* bajo la ayuda post-doctoral APOSTD/2016/042

[2] Si el humano seleccionara la opción de modo manual se seguiría la plantilla de la tarea "Tomar control del coche"

## Referencias


1. Cámara, J., Moreno, G., Garlan, D.: Reasoning about human participation in self- adaptive systems. In: SEAMS 2015, pp. 146–156 (2015)
2. Dorn, C., Taylor, R.N.: Coupling software architecture and human architecture for collaboration-aware system adaptation. In: ICSE, pp. 53–62 (2013)
3. ERTRAC. Automated Driving Roadmap. ERTRAC Task Force. Connectivity and Automated Driving, (2015).
4. Evers, C., Kniewel, R., Geihs, K., Schmidt, L.: The user in the loop: enabling user participation for self-adaptive applications. FGCS J. 34, pp. 110–123 (2014)
5. Farooq, U., Grudin, J.: Human-computer integration**.** Interactions 23(6): pp. 26-32 (2016)
6. Fitts, P. M.: Human engineering for an effective air-navigation and traffic-control system. Washington, DC: National Research Council, (1951).
7. Gil, M., Pelechano, V.: Self-adaptive unobtrusive interactions of mobile computing systems. JAISE 9(6): 659-688 (2017)
8. Laddaga, R.: Active software. Int. Workshop on Self-Adaptive Software. pp. 11-26 (2000)
9. Lee, J. D., See, K. A.: Trust in automation: Designing for appropriate reliance. Human Factors, 46, pp. 50–80, (2004).
10. Litman, T.: Autonomous Vehicle Implementation Predictions. Implications for Transport Planning. Victoria Transport Policy Institute. Febrero 2018
11. Moore, A., O'Reilly, T., Nielsen, P. D., Fall, K.: Four Thought Leaders on Where the Industry Is Headed. IEEE Softw. 33, 1, pp. 36-39 (2016).
12. Nunes, D. S., Zhang, P., Sá Silva, J.: A Survey on Human-in-the-Loop Applications Towards an Internet of All. IEEE Communications Surveys & Tutorials, 17(2), pp. 944-965 (2015).
13. Parasuraman, R., Sheridan, T. B., Wickens, C. D. 2000.: A model for types and levels of human interaction with automation. Trans. Sys. Man Cyber. Part A 30, 3, pp. 286-297, (2000).
14. Pelechano, V., Gil, M., Fons, J., Albert, M.: Diseñando la Participación del Humano en los Sistemas Autónomos. JISBD 2017, pp. 1-14 (2017)
15. Row, S.: The Future of Transportation: Connected Vehicles to Driverless Vehicles...What Does It Mean To Me?, ITE Journal, Vol. 83, No. 10, pp. 24-25, (2013).
16. Salehie, M., Tahvildari, L.: Self-adaptive software: Landscape and research challenges. Journal ACM Transactions on Autonomous and Adaptive Systems Vol. 4 No. 2, (May 2009)
17. SAE (Society of Automotive Engineers) International. Surface Vehicle Recommended Practice. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. J3016, septiembre 2016
18. Sheridan, T. B., Verplank, W. L.: Human and Computer Control of Undersea Teleoperators," MIT Man-Machine Systems Laboratory, MA, Tech. Rep., (1978).
19. Shin, C., Dey, A.K., Woo, W.: Mixed-initiative conflict resolution for context-aware applications. In: UbiComp 2008, pp. 262–271 (2008)
20. Stanton, N. A., Young, M. S.: Vehicle automation and driving performance. Ergonomics, 41(7), pp. 1014-1028, (2010).
21. Trimble, T. E., Bishop, R., Morgan, J. F., Blanco, M.: Human factors evaluation of level 2 and level 3 automated driving concepts: Past research, state of automation technology, and emerging system concepts. (Rep. No. DOT HS 812 043), (2014, July).
22. Russell, D. M., Maglio, P., Dordick, R., and Neti, C. 2003. Dealing with ghosts: Managing the user experience of autonomic computing. IBM Sys. Journal 42, 1, 177-188.
23. Stumpf, S., Burnett, M., Pipek, V., Wong, W.K. End-user interactions with intelligent and autonomous systems. In CHI EA'12 Conference, pp. 2755-2758 (2012).